

UNIVERSIDADE FEDERAL DO PARANÁ

Helba Cirino de Souza Barbosa

**Re-Anotação do Genoma de *Herbaspirillum seropedicae* SmR1 com Dados de
Transcriptoma por RNA-Seq**

CURITIBA

2014

UNIVERSIDADE FEDERAL DO PARANÁ

Helba Cirino de Souza Barbosa

**RE-ANOTAÇÃO DO GENOMA DE *HERBASPIRILLUM SEROPEDICAE*
SmR1 COM DADOS DE TRANSCRIPTOMA POR RNA-Seq**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Bioinformática, Programa de Pós-Graduação em Bioinformática, Setor de Educação Profissional e Tecnológica da Universidade Federal do Paraná.

Orientadora: Prof. Dr^a Rose Adele Monteiro

Co-orientador: Prof. Dr. Leonardo Magalhães Cruz

Co-orientador: Prof. Dr. Emanuel Maltempi de Souza

CURITIBA
2014

TERMO DE APROVAÇÃO

HELBA CIRINO DE SOUZA BARBOSA

“Re-anotação do genoma de *Herbaspirillum seropedicae* SmR1 baseada em dados de transcriptoma por RNA-Seq”

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Bioinformática, pelo Programa de Pós-graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná, pela seguinte banca examinadora:


Orientadora:

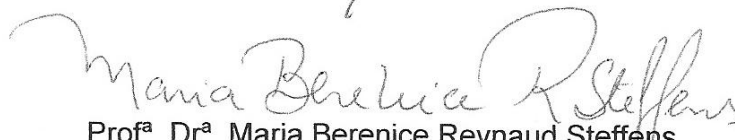

Prof^a Dr^a Rose Adele Monteiro


Coorientador:


Prof. Dr. Emanuel Maltempi de Souza

Coorientador:


Prof. Dr. Leonardo Magalhães Cruz


Prof^a Dr^a Maria Berenice Reynaud Steffens
Universidade Federal do Paraná


Prof. Dr. Joseph Albert Medeiros Evaristo
Universidade Federal do Paraná

Curitiba, 02 de abril de 2014

AGRADECIMENTOS

Uma conquista não é uma única conquista, vem recheada de muitas outras. É assim que me sinto. Quantas realizações consegui ao buscar o mestrado em Bioinformática! Quantas transformações internas, quantos amigos, quantas superações, quanto aprendizado!

Meu coração se enche de alegria ao ver tanta beleza que há em querer, em buscar e em conseguir. E principalmente em me transformar, abrir portas dentro de mim para a compreensão se expandir.

Hoje vejo que tudo está em seu devido lugar, que tudo aconteceu em direção ao meu crescimento. E agradeço:

- Ao Omar Cirino de Souza (papai), à Edna Sousa Barbosa (mamãe) e ao Rogério Sousa Dantas (irmão) por sempre terem me incentivado nos estudos, dando todo o apoio necessário;

- Aos professores Dra. Rose Adele Monteiro, Dr. Leonardo Magalhães Cruz e Dr. Emanuel Maltempi de Souza pela oportunidade e ensinamentos;

- Às professoras Dra. Maria Berenice Reynaud Steffens, Dra. Jeroniza Nunes Marchaukoski e aos professores Dr. Roberto Tadeu Raittz e Dr. Adriano Barbosa Silva;

- À Dra. Michele Tadra, ao Me. Rodrigo Luiz Cardoso, ao Dr. Helisson Faoro, ao Dr. Daniel Renato Lammel, ao Dr. Joseph Albert Medeiros Evaristo, ao Dr. Vinicius Weiss, ao Me. Leandro Moreno;

- Aos colegas do mestrado em Bioinformática das turmas de 2011, 2012, 2013 e à Léa.

- À Susana Gobetti.

- Ao Dr. Issakar Lima Souza (UFSC) e à Dra. Ludier Kesser Santos Silva (UFMT) por me incentivarem a cursar mestrado.

- Às amigas Mariana Kirsten, Ananda Flor F. Boechat e Aila Zittlau, por se fazerem presentes em minha vida em todos momentos, mesmo distantes;

- Aos amigos-irmãos CEBUDV e à família do Clayton Gomes e Rutneia;

- Ao meu namorado Bernardo Fermino Correia de Lima, por fazer dessa fase de mestrado dias agradáveis, pelo apoio e incentivo no dia-a-dia.

RESUMO

As plantas necessitam do Nitrogênio para sintetizar várias moléculas que são essenciais para sua sobrevivência, porém, não conseguem fixá-lo a partir do ar atmosférico. É neste contexto que entram em cena as bactérias diazotróficas, que interagem com as plantas e fixam o nitrogênio atmosférico convertendo-o a uma forma que pode ser assimilada pelas plantas, contribuindo para o crescimento vegetal. Tendo em vista a importância destas bactérias para o desenvolvimento da agricultura, em 2011, foi sequenciado o genoma completo da bactéria diazotrófica *Herbaspirillum seropedicae* SmR1 pelo Programa Genoma do Paraná (Genopar). Posteriormente, foram obtidos dados do transcriptoma desta bactéria crescida em diferentes condições ambientais: NFbHPN-Malato, NFbHPN-Malato-Naringenina, NFbHPN-Malato-Nitrato, Planctônico-milho-1dia, Planctônico-milho-3dias, Aderido-milho- 1dia, Aderido-milho3dias, Planctônico- trigo, Aderido-trigo, NFbHPN-Malato-Alto Oxigênio, NFbHPN-Malato-Baixo Oxigênio. A análise dos transcriptomas com base nos dados de RNA-Seq permitiu a identificação de novos genes, a localização dos genes essenciais e de operons, bem como de ncRNAs. A identificação de novos genes permite a reanotação do genoma de *H. seropedicae* com base nos dados de RNA-Seq.

Palavras-chaves: *Herbaspirillum seropedicae*, diazotrófica, transcriptoma, RNA-Seq.

ABSTRACT

Plants need nitrogen to synthesize various molecules that are essential for their survival, however, they can not fix it from atmospheric air. Diazotrophic bacteria, can fix atmospheric nitrogen by converting it to a form that can be assimilated by plants, contributing to plant growth. In 2011, the genome of the diazotrophic bacteria called *Herbaspirillum seropedicae* SMR1 was sequenced by Genome Program of Paraná (Genopar). Subsequently, we obtained data from the transcriptome of this bacterium under different environmental conditions: NFbHPN-Malate, NFbHPN-Malate-Naringenin, NFbHPN-Malate-Nitrate, Planktonic-maize-1day, Planktonic-maize-3 days, maize root attached 1 day-, maize root attached 3 days, Planktonic-wheat, wheat root attached, NFbHPN-Malate-High Oxygen, NFbHPN-Malate-Low Oxygen. The analysis of the transcriptome data based on RNA Seq allowed the identification of new genes, the location of the essential genes and operons, as well as non-coding RNA. The identification of new genes permit the reannotation of the *H. seropedicae* genome of based on RNA - Seq data.

Keywords: *Herbaspirillum seropedicae*, diazotrophic, transcriptome, RNA - Seq.

LISTA DE ILUSTRAÇÕES

Figura 01: bactéria <i>Herbaspirillum seropedicae</i> com três flagelos	12
Figura 02: genoma de <i>Herbaspirillum seropedicae</i> SmR1	13
Figura 03: identificação de estruturas de operons e de um novo gene a partir de dados de transcriptoma	16
Figura 04: sequência de trabalho do Rockhopper	26
Figura 05: gráfico tipo BoxWhisker dos dados ‘brutos’ do sequenciamento de uma amostra da condição Nitrato	27
Figura 06: gráfico tipo BoxWhisker dos dados após trinar as últimas 10 bases e realizar o filtro de qualidade Q20 do sequenciamento de uma amostra da condição Nitrato	28
Figura 07: fluxo de tratamento dos dados de RNA-Seq	29
Figura 08: classificação funcional realizada pelo COG	35
Figura 09: classificação funcional realizada pelo COG dos 50 genes mais expressos em todas as condições analisadas	36
Figura 10: Proteínas ribossomais de <i>H. seropedicae</i>	38
Figura 11: a via metabólica ciclo de Krebs	39
Figura 12: a via metabólica metabolismo do piruvato	40
Figura 13: via metabólica degradação de RNA	40
Figura 14: via metabólica montagem flagelar	41
Figura 15: via metabólica da glicose	42
Figura 16: via metabólica fosforilação oxidativa	43
Figura 17: gráfico de colunas dos índices CAI, CBI, Fop e GC3	45

Figura 18: visualização no Artemis da expressão do operons *nuoA-nuoD* na condição

NFbHPN-Malato-Naringenina 47

LISTA DE TABELAS

Tabela 01: características do genoma de <i>H. seropedicae</i>	13
Tabela 02: parâmetros para mapeamento dos <i>reads</i> com o CLC Workbench	29
Tabela 03: tamanhos das bibliotecas	30
Tabela 04: relação dos 50 genes mais expressos nas onze condições de transcriptoma, com seus respectivos produtos e vias metabólicas das quais participam ou função	31
Tabela 05: média e erro padrão da média dos índices CAI, CBI, Fop e GC3 dos genes constitutivos	44
Tabela 06: média e erro padrão da média dos índices CAI, CBI, Fop e GC3 dos genes não constitutivos	44
Tabela 07: operons para os genes NuoA-nuoN	47
Tabela 08: quantidade de novos genes hipotéticos por condição	48
Tabela 09: funções gênicas, de 18 novos genes que tiveram alguma função retornada pelo KEGG	49
Tabela 10: quantidade de ncRNAs confirmados em cada condição	52

SUMÁRIO

1. Introdução.....	12
1.1 <i>Herbaspirillum seropedicae</i>	12
1.2. Transcriptoma do <i>H. seropedicae</i> em diferentes condições de cultivo	14
2. Objetivo Geral	17
2.1 Objetivos Específicos	17
3. Justificativa.....	17
4. Materiais e Métodos.....	18
4.1 Condições dos Transcriptomas	18
4.2 RPKM e Cobertura.....	20
4.3 <i>Softwares</i>	21
4.3.1 Artemis	21
4.3.2 CLC DNA Workbench	21
4.3.3 Samtools.....	22
4.3.4 CodonW	22
4.3.5 SHRiMP	23
4.3.6 KEGG e KAAS	23
4.3.7 WampServer.....	23
4.3.8 Quality Assessment	24
4.3.9 GraphPad Prism 5	24
4.3.10 FastQC	24
4.3.11 Preditor de operons Rockhopper	25
5. Resultados e Discussão	27
5.1 Análise de qualidade	27
5.2 Identificação dos genes essenciais de <i>H. seropedicae</i> utilizando os dados de transcriptoma de onze condições de cultivo diferentes.....	28

5.3 Classificação Funcional das proteínas codificadas pelos genes identificados	35
5.4 Distribuição das proteínas codificadas pelos genes identificados nas principais vias metabólicas	37
5.5 Análise dos 538 genes identificados com o software CodonW	44
5.6 Identificação dos operons presentes em <i>H. seropedicae</i> utilizando dados de transcriptoma	46
5.7 Identificação de genes de <i>H. seropedicae</i> ainda não anotados utilizando dados de RNAseq.....	47
5.7.1 HGF	48
5.7.2 Identificação de ncRNA de <i>H. seropedicae</i>	50
ANEXO 1 - Lista dos genes expressos nas 11 condições de transcriptoma com RPKM ≥ 100 OU Cobertura ≥ 2	61
ANEXO 2 - 861 operons retornados pelo <i>software</i> Rockhopper.....	64
ANEXO 3 - Tabela dos genes preditos pelo <i>software</i> HGF e expressos com cobertura ≥ 3	86
ANEXO 4 - tabela dos ncRNA preditos por Moreno (2013) e expressos com cobertura ≥ 5	96
ANEXO 5: <i>Script</i> em Perl para trimar os arquivos de qualidade	99
ANEXO 6: <i>Script</i> em Perl para trimar os arquivos csfasta.	100

1. Introdução

1.1 *Herbaspirillum seropedicae*

O nitrogênio (N) é um elemento essencial para a síntese de biomoléculas, porém a maior parte do N, não é assimilável para a maioria dos organismos, pois está presente na forma de nitrogênio atmosférico (N₂). Algumas bactérias e arqueas possuem a capacidade de reduzir N₂ a NH₃, processo denominado fixação biológica de nitrogênio (FBN). Bactérias fixadoras de nitrogênio, chamadas de diazotróficas, estão presentes nas plantas e são capazes de transferir o nitrogênio fixado para a planta (DÖBEREINER & PEDROSA, 1987).



FIGURA 01. Bactéria *Herbaspirillum seropedicae* com três flagelos.
FONTE: <<http://cienciahoje.uol.com.br/blogues/bussola/2011/08/descricao-detalhada>> acessado em 28/08/2012.

O *Herbaspirillum seropedicae* é uma Betaproteobacteria, fixadora de nitrogênio, isolada de milho (*Zea mays*), sorgo (*Sorghum bicolor*) e arroz (*Oryza sativa*) (BALDANI *et al.*, 1986). O *H. seropedicae* coloniza várias plantas hospedeiras. A associação com essas plantas geralmente começa com a atração e a adesão da bactéria a raiz, colonização e invasão dos pontos de emergência das raízes laterais e ocupação dos espaços intercelulares e xilema (JAMES & OLIVARES, 1998). O *H. seropedicae* é uma bactéria promotora do crescimento vegetal que quando inoculado causou um aumento total na biomassa de arroz e cana de açúcar (BODDEY *et al.* 1995; BALDANI *et al.* 2000; JAMES *et al.* 2002; GYANESHWAR *et al.* 2002). Esse efeito pode ser devido a fatores como: fixação biológica de nitrogênio, produção de fitohormônios e sideróforos e atividade de ACC deaminase (MONTEIRO *et al.*, 2012).

Devido à importância do *H. seropedicae* como biofertilizante, vários estudos tem sido realizados com essa bactéria. O genoma do *H. seropedicae* SmR1 foi sequenciado e anotado pelo Programa Genoma do Paraná (Genopar) (PEDROSA *et al.*, 2011). As características principais do genoma de *H. seropedicae* estão organizadas na Tabela 01 e na Figura 02.

TABELA 01: características do genoma de *H. seropedicae*. FONTE: PEDROSA *et al.* (2011).

Características gerais do genoma de <i>Herbaspirillum seropedicae</i> SmR1	
Tamanho	5.513.887 bp
GC	63,4%
Total de genes	4.804
Total de CDS	4.735
Tamanho médio das ORF	1.028bp
Regiões codificadoras de proteínas	88,3%
rRNA operons	3
tRNAs	55
Genes com atribuição funcional	3.108
Predição função geral	497
Função desconhecida	1.130
Supostas regiões de transferência horizontal	18

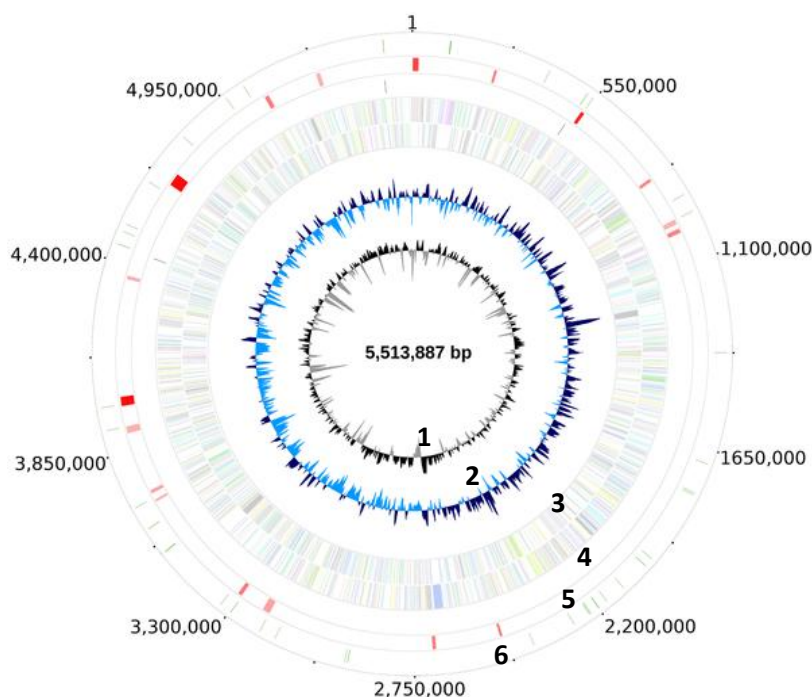


FIGURA 02. Genoma de *Herbaspirillum seropedicae*, estirpe SmR1. 1) conteúdo de G+C. 2) correlação de G, GC: GC skew. 3) código de cores dos genes de acordo com sua classificação funcional. 4) operons de rRNAs. 5) regiões de supostas transferência horizontal: vermelho claro indica baixa pontuação e vermelho escuro indica alta pontuação. 6) regiões de *H. seropedicae* idêntica a mamona (*Ricinus communis*) FONTE: PEDROSA *et al.* (2011).

O sequenciamento do genoma permitiu a identificação dos genes presentes nessa bactéria e os produtos que são codificados por eles. Essas informações permitiram o mapeamento das vias metabólicas e dos processos que esta bactéria é capaz de realizar. Foram encontrados genes envolvidos com os fatores que podem estar relacionados com a promoção do crescimento vegetal: genes envolvidos com a interação planta:bactéria, fixação biológica de nitrogênio, produção de fitohormônios e sideróforos e atividade de ACC deaminase (PEDROSA *et al.*, 2011). Para verificar quando esses genes são expressos, foram feitos vários experimentos de transcriptomas com o *H. seropedicae* sendo cultivado em diferentes condições. A união dos dados de genoma e transcriptoma permite desvendar melhor o comportamento desta bactéria em diferentes condições de cultivo.

1.2. Transcriptoma do *H. seropedicae* em diferentes condições de cultivo

O conjunto de todos os genes expressos por células, tecidos, ou organismos em um determinado espaço e tempo é denominado transcriptoma. Inclui todos os tipos de RNA: mRNA, tRNA, rRNA, miRNA, ncRNA dentre outros. Esse ramo da genômica funcional permite investigar o padrão de expressão de milhares de genes simultaneamente, de um organismo ou tipo celular, em um habitat ou condição em uma determinada fase da vida (MARQUES & SILVA, 2004).

Os dados utilizados neste trabalho foram obtidos pela técnica de RNA-Seq, uma tecnologia de sequenciamento em larga escala que pode revelar a maioria dos tipos de RNA dentro de uma célula, fornecendo dezenas a centenas de milhões de sequências e informações sobre bilhões de bases individuais. No entanto, esta técnica exige investir tempo para desenvolver uma compreensão sofisticada dos dados usando bioinformática. É fácil de iniciar um projeto, mas é difícil de obter e interpretar dados para responder adequadamente as perguntas experimentais (CHU & COREY, 2012).

Neste trabalho foram utilizados os dados de RNA-Seq de 11 condições de transcriptoma de *H. seropedicae*, conforme descritas na tabela 02.

Na condição NFbHPN-Malato houve 3711 genes expressos e na condição NFbHPN-Malato-Naringenina houve 3672 genes expressos. Ao fazer uma análise comparativa dessas duas condições percebeu-se 336 genes diferencialmente expressos, permitindo a visualização de como acontece a transcrição dos genes na presença de naringenina (TADRA-SFEIR, 2011).

Com a análise do transcriptoma na condição NFbHPN-Malato-Nitrato foi possível entender melhor a regulação de genes envolvidos com o metabolismo do nitrato e os papéis dos genes que codificam para as prováveis enzimas nitrato redutases assimilatória e respiratória (BONATO, 2012).

A análise das condições Planctônico–milho-1dia, Planctônico–milho-3dias, Aderido–milho- 1dia, Aderido–milho- 3dias, Planctônico- trigo e Aderido-trigo promoveram esclarecimentos sobre aspectos da interação planta-bactéria (BALSANELLI, 2013; PANKIEVICZ, 2013).

As condições NFbHPN-Malato-Alto Oxigênio e NFbHPN-Malato-Baixo Oxigênio de *H. seropedicae* e de mutantes sugerem que proteínas Fnr são os participantes principais na regulação da composição da cadeia de transporte de elétrons em resposta a concentrações de oxigênio (BATISTA, 2013).

Dados de transcriptoma de *Salmonella typhi* indicaram haver expressão de quarenta e dois RNAs não codificadores de proteínas (PERKINS, 2009). E sessenta e sete ncRNAs em *L. monocytogenes* (OLIVER *et al.*, 2009).

Genes em genomas procarióticos são muitas vezes dispostos em operons, com um RNAm policistrônico abrangendo vários co-transcritos de genes. A predição de estruturas de operons é geralmente baseada na ocorrência de uma curta distância entre genes consecutivos, a conservação de ordem de genes em organismos relacionados, e a identificação das associações funcionais entre os genes adjacentes. Várias ferramentas de previsão de operons estão disponíveis online, mas estas possuem uma precisão limitada (SOREK & COSSART, 2010).

Recentemente vem surgindo *softwares* que utilizam dados de RNA-Seq para prever operons, com resultados promissores (LI *et al.*, 2014 ; MCCLURE *et al.*, 2013).

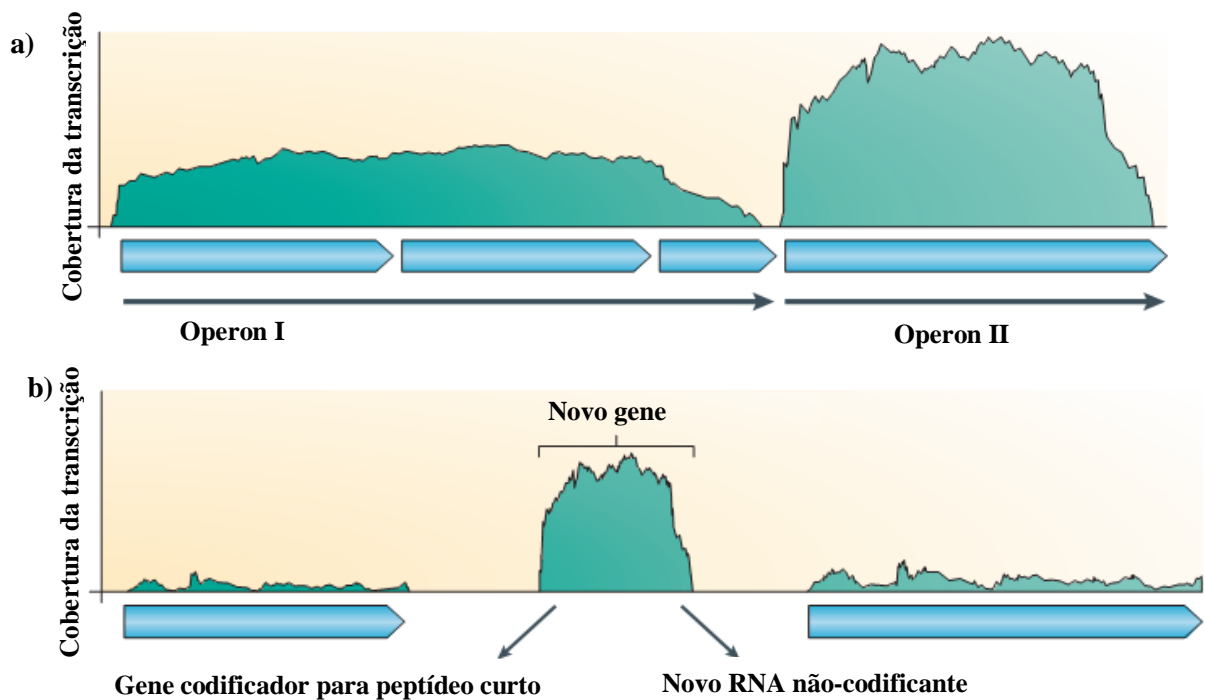


FIGURA 03. a) Identificação de estruturas de operons com dados de transcriptoma. b) Identificação de um novo gene a partir de dados de transcriptoma. FONTE: SOREK & COSSART (2010).

O genoma da bactéria *H. seropedicae* foi anotado por *software* de previsão de gene e comparações com *softwares* Blast que identificaram genes codificadores de proteínas (PEDROSA *et al.*, 2011). Não foram detectados alguns genes de pequeno porte ou genes que ainda não estavam anotados em outros organismos e os RNAs não codificadores. A análise da cobertura do transcriptoma permite a identificação dessas regiões transcritas (SOREK & COSSART, 2010).

Os dados de RNA-Seq de 11 condições de crescimento de *H. seropedicae* foram utilizados para identificar genes essenciais, genes ainda não anotados (codificadores de proteínas ou não) e operons (Figura 03).

2. Objetivo Geral

- Reanotar o genoma de *Herbaspirillum seropedicae* SmR1 com dados de transcriptoma obtidos por RNA-seq nas seguintes condições: NFbHPN-Malato, NFbHPN-Malato-Naringenina, NFbHPN-Malato-Nitrato, Planctônico–milho-1dia, Planctônico-milho-3dias, Aderido–milho- 1dia, Aderido-milho3dias, Planctônico- trigo, Aderido-trigo, NFbHPN-Malato-Alto Oxigênio, NFbHPN-Malato-Baixo Oxigênio.

2.1 Objetivos Específicos

- Mapear todo conjunto de dados de RNAseq de *H. seropedicae* SmR1;
- Analisar proteínas hipotéticas e/ou não anotadas;
- Identificar os genes encontrados em operons;
- Analisar os genes anotados pelo HGF - Hybrid Gene Finder, (RAITTZ, 2010);
- Identificar genes que são sempre expressos (genes essenciais) em 11 condições de crescimento diferentes por RNA-seq;
- Validar ncRNA preditos.

3. Justificativa

O *H. seropedicae* é uma bactéria promotora de crescimento vegetal, cujo uso pode diminuir a utilização de fertilizantes nitrogenados, reduzindo o custo de produção, a poluição do lençol freático e outros danos causados pela utilização de fertilizantes nitrogenados.

O estudo da expressão gênica de *H. seropedicae* em diferentes condições de cultivo, inclusive durante a interação com a planta, permitirá estabelecermos quais são os produtos gênicos essenciais e quais são os importantes para cada condição. Os dados de RNA-Seq também permitirão a anotação de genes que estão sendo expressos e não estão anotados, quer

seja genes codificadores de proteína ou não codificador de proteína (ncRNA). O refinamento da anotação genômica através da análise de expressão por RNA-Seq permite obter uma maior precisão na identificação dessas regiões funcionais.

4. Materiais e Métodos

4.1 Condições dos Transcriptomas

Para realizarmos a reanotação do genoma de *H. seropedicae* determinamos o transcriptoma desta bactéria em condições distintas, descritas no quadro seguinte:

QUADRO 01. Descrição das condições de cultivo bacteriano para obtenção dos transcriptomas obtidos com a tecnologia SOLiD.

Condições	Descrição
1. NFbHPN-Malato	A condição controle consiste nas bactérias cultivadas a 30°C em meio NFbHPN, utilizando malato como fonte de carbono até D.O 0,8 (TADRA-SFEIR, 2011).
2. NFbHPN-Malato-Naringenina	Bactérias cultivadas a 30°C em meio NFbHPN, utilizando malato como fonte de carbono na presença de 100µM do flavonóide naringenina, até D.O. 0,8. (TADRA-SFEIR, 2011).
3. NFbHPN-Malato-Nitrato	Bactérias cultivadas a 30°C em meio NFbHPN, utilizando malato como fonte de Carbono e nitrato 10mM como fonte de nitrogênio. (BONATO, 2012).
4. Planctônico–milho-1 dia	Bactérias recuperadas do meio líquido do sistema hidropônico (denominadas

Condições	Descrição
	plantônicas) após um dia de inoculação em raiz de milho (BALSANELLI, 2013).
5. Planctônico-milho-3dias	Bactérias recuperadas do meio líquido do sistema hidropônico (denominadas plantônicas) após um dia de inoculação em raiz de milho (BALSANELLI, 2013).
6. Aderido-milho- 1dia	Bactérias recuperadas da superfície radicular de milho (denominadas aderidas) após um dia de inoculação (BALSANELLI, 2013).
7. Aderido-milho3dias	Bactérias recuperadas da superfície radicular de milho (denominadas aderidas) após três dias de inoculação (BALSANELLI, 2013).
8. Planctônico- trigo	Bactérias recuperadas do meio líquido do sistema hidropônico (denominadas plantônicas) após três dias de inoculação em raiz de trigo (PANKIEVICZ, 2013).
9. Aderido-trigo	Bactérias recuperadas da superfície radicular de trigo (denominadas aderidas) após três dias de inoculação (PANKIEVICZ, 2013).
10. NFbHPN-Malato-Alto Oxigênio	Bactérias cultivadas a 30°C em meio NFbHPN, utilizando malato como fonte de carbono até D.O. de 0,4. Depois de atingida a D.O. 0,4 as bactérias continuaram sendo cultivadas nesta condição por mais 1,5 horas. (BATISTA, 2013).
11. NFbHPN-Malato-Baixo Oxigênio	Bactérias cultivadas a 30°C em meio NFbHPN, utilizando malato como fonte de

Condições	Descrição
	carbono até D.O. de 0,4. Nesse momento as bactérias passaram a ser cultivadas em condições microanaerobias (concentração inicial de oxigênio 2%) por 1,5 horas. (BATISTA, 2013).

Os RNAs ribossomais 16S e 23S foram excluídos, pois não são traduzidos e por serem muitos (aproximadamente 90% dos transcritos) tornam o processo de reanotação demorado. Para cada condição de cultivo foi utilizada duas replicatas.

4.2 RPKM e Cobertura

Para calcular a expressão de um gene é necessária uma normalização da quantidade de leituras para permitir a comparação intra e entre as bibliotecas de transcritos. A normalização foi feita pelo método de RPKM (*reads per kilobase of transcript per million mapped reads*) (MORTAZAVI & WILLIAMS *et al.*, 2008), que mensura o valor de expressão de um gene levando em consideração o tamanho da biblioteca, o tamanho dos genes e o número de leituras por gene, utilizando a fórmula: $RPKM = (10^9 C) / (NL)$, onde C = número de leituras por gene, L = tamanho do gene, N = tamanho da biblioteca (número total de leituras mapeadas).

A cobertura estima quantas vezes a quantidade de leituras pode cobrir o tamanho do gene. Cobertura = $(C * N) / L$, onde C = número de leituras por gene, L = tamanho do gene e N é o tamanho das leituras.

4.3 Softwares

Os *softwares* descritos abaixo foram utilizados para a análise comparativa do transcriptoma do *Herbaspirillum seropedicae* em condições distintas.

4.3.1 Artemis

O *software* Artemis é um navegador de genoma gratuito e uma ferramenta de anotação que permite visualizar as características da sequência, dados de sequenciamento feitos em plataformas de nova geração e os resultados das análises de acordo com o contexto da sequência, e também a sua tradução nas fases de leitura (*frames*) possíveis. Pode ler arquivos nos formatos EMBL, GenBank ou sequência em FASTA, FASTA indexados, GFF ou formato cru (RUTHERFORD, 2000). Podendo ser utilizado para visualização do genoma de *Herbaspirillum seropedicae* e realizar as análises necessárias.

Artemis é escrito em Java e está disponível para as plataformas UNIX, Macintosh e Windows. É um *software* livre distribuído sob os termos da Licença Pública Geral (GPL/GNU). Deve funcionar em qualquer sistema com uma versão recente do Java (SANGER, 2012).

4.3.2 CLC DNA Workbench

O CLC DNA *Workbench* é uma bancada de trabalho ‘virtual’ que possibilita análises em sequências de DNA, RNA e proteínas. O *software* suporta ler mapeamento de Sanger e dados de sequenciamento das tecnologias SOLiD, Ion Torrent, 454 e Illumina; montagem de novo, detecção de SNP em genomas inteiros, a detecção de variações estruturais dos genomas, análise de Chip-seq, RNA-Seq, análise de *small*RNA e BLAST. Permite analisar, visualizar e comparar dados genômicos, transcriptomas e epigenômicos de todas as principais plataformas de *high-throughput*. Além disso, é independente de sistema operacional (CLC BIO, 2012).

4.3.3 Samtools

SamTools fornece vários utilitários para manipulação de alinhamentos no formato SAM, incluindo separar, mesclar, indexar, converter em outros formatos e gerar alinhamentos em um formato por posição (SAMTOOLS, 2012; LI, 2009).

4.3.4 CodonW

CodonW é um programa desenvolvido para simplificar a análise multivariada (análise de correspondência) de códon e uso de aminoácidos. Ele também calcula índices de padrão de utilização de códons. Tem interface de menu e de linha de comando (CODONW, 2013).

O software CodonW realiza cálculos dos índices CAI, CBI, Fop e GC3s:

CAI: (Codon Adaptation Index – índice de adaptação de códon) é uma medida da capacidade de adaptação da utilização do codon de um gene para a utilização de códons de genes altamente expressos. A adaptabilidade relativa de cada codon é a razão entre o uso de cada codon e o codon mais abundante para o mesmo aminoácido (SHARP & LI, 1987).

CBI: (Codon Bias Index – Índice de Tendência de Códon) mede o grau em que um gene utiliza um subconjunto de códons ideais. Em um gene com alta tendência de uso de códons, CBI será igual a um, um gene com a utilização de códons aleatórios CBI será igual à zero. É possível que o número de códons ideais seja menor do que o esperado pela mudança aleatória. Isso resulta em um valor negativo para CBI (BENNETZEN & HALL, 1982).

Fop: (Frequency of Optimal codons – Frequência de Códon Ótimos) é a razão entre os códons ótimos para os códons sinônimos (código genético dependente). Por padrão, os códons ideais de *E. coli* são assumidos. O usuário também pode escolher os códons ideais. Os valores de Fop para o índice original são sempre entre zero (quando códons ideais não são usados) e um (onde apenas códons ideais são usados) (IKEMURA, 1981).

GC3s: fração de códons, que tenham guanina ou citosina na terceira posição do códon.

4.3.5 SHRiMP

SHRiMP é um pacote de *softwares* para alinhamento de *reads* contra o genoma alvo. Ele foi desenvolvido primeiramente tendo em mente os numerosos *reads* de máquinas de sequenciamento de próxima geração. O método baseia-se em uma técnica de mapeamento rápido dos *reads* (RUMBLE *et al*, 2009).

4.3.6 KEGG e KAAS

O KEGG (*Kyoto Encyclopedia of Genes and Genomes* - Kyoto Enciclopédia de Genes e Genomas – KEEG, 2013) é um recurso de banco de dados para a compreensão das funções de sistemas biológicos, como a célula, o organismo e o ecossistema, a partir de informações em nível molecular, especialmente em conjuntos de dados moleculares gerados pelo sequenciamento do genoma e outras tecnologias experimentais de elevado rendimento.

O KAAS (KEGG *Automatic Annotation Server* - KEGG Servidor de anotação automática) fornece anotação funcional de genes por meio de comparações BLAST contra o banco de dados manualmente curado GENES KEGG. O resultado contém KO (KEGG *Orthology*) que é um identificador automatizado de atribuições e vias metabólicas (MORIYA *et al.*, 2007 ; KANEHISA *et al.*, 2013).

4.3.7 WampServer

WAMP é um acrônimo para Windows, Apache, MySQL e PHP, que são linguagens e ferramentas para desenvolvimento web do Windows. É uma plataforma que permite o desenvolvimento de aplicações web e um banco de dados MySQL. O phpMyAdmin permite a gerência do bancos de dados.

4.3.8 Quality Assessment

Com o *software* Quality Assessment pode-se obter e analisar gráficos que mostram a distribuição dos valores de qualidade a partir do sequenciamento. Este software permite adotar as normas de qualidade mais rigorosas para os dados de sequência, com base na análise de qualidade de gráficos e cobertura estimada após a aplicação do filtro de qualidade, proporcionando uma sequência aceitável para as inferências biológicas. O filtro da qualidade é um passo fundamental no processo de montagem de transcriptomas, uma vez que reduz a frequência de alinhamentos incorretos que são causados por erros de precisão, devido ao tamanho curto das leituras (50 pb). Aplicação de filtros de qualidade utilizando o *software* Quality Assessment, proporciona uma maior precisão na definição dos parâmetros de corte e no transcriptoma como um todo (RAMOS *et al*, 2011).

4.3.9 GraphPad Prism 5

O *software* GraphPadPrism foi utilizado para as análises estatísticas do conjunto de dados dos genes constitutivos e não constitutivos. Os índices CAI, CBI, Fop e GC3 (descritos no item 4.3.4) foram analisados pelos teste *t* de Student e expressos como média±erro padrão da média.

4.3.10 FastQC

Sequenciadores em larga escala podem gerar dezenas de milhões de sequências em uma única corrida. Antes de analisar essas sequências para tirar conclusões biológicas é necessário realizar alguns testes de controle de qualidade para garantir que os dados brutos sejam confiáveis e que não há problemas nos dados que podem afetar a interpretação (FASTQC, 2012).

O FastQC está disponível em: www.bioinformatics.babraham.ac.uk/projects/fastqc/

4.3.11 Preditor de operons Rockhopper

O Rockhopper foi desenvolvido em Java, sendo um sistema abrangente e de interface amigável para a análise computacional de dados de RNA-seq bacteriano. Rockhopper faz *upload* de arquivo(s) de reads de sequenciamento em larga escala nos formatos Fastq, Qseq, Fasta, Sam, ou Bam.

Rockhopper realiza as seguintes tarefas: alinhamento de *reads* a um genoma, normalização de dados de diferentes experimentos, montagem de transcriptoma, identifica os limites de transcrição e novas transcrições, como RNAs curtos, quantifica expressão, expressão diferencial de genes, localização de operons e visualização dos resultados em um navegador de genoma (MCCLURE *et al.*, 2013). As etapas do programa estão representadas na figura 07. E está disponível no endereço <http://cs.wellesley.edu/~btjaden/Rockhopper/>.

No programa são utilizadas duas funcionalidades para estimar a probabilidade de genes serem operons: a distância entre os genes e a similaridade de expressão dos genes nos dados de RNA-seq. Os recursos são combinados, junto com a probabilidade de ser um operon, usando um classificador Naive Bayes com operons prévios para determinar a probabilidade de serem co-transcritos. O desempenho do sistema foi avaliado com os dados obtidos de RNA-Seq realizados com cinco bactérias diferentes (MCCLURE *et al.*, 2013).

O programa foi utilizado com suas configurações e parâmetros padrões para prever operons com base em dados de RNA-Seq em 11 condições de cultivo da bactéria *H. seropedicae*.

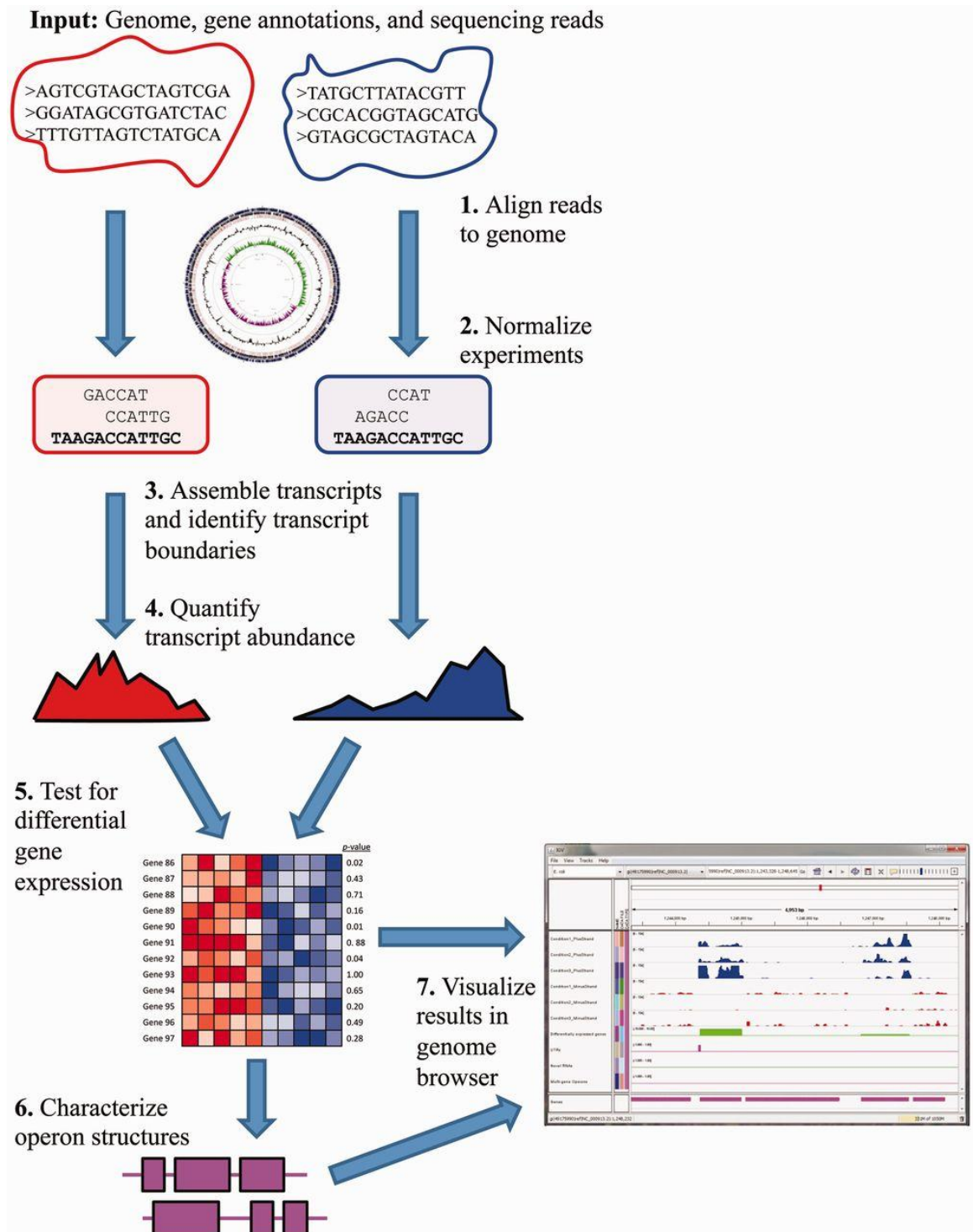


FIGURA 04. Sequência de trabalho do Rockhopper. A entrada para Rockhopper consiste de uma sequência do genoma (arquivo FASTA) e as anotações de genes (arquivos PPF e RNT) e *reads* sequenciados (FASTQ, QSEQ ou FASTA). Os resultados do programa saem como arquivos de texto delimitados por tabulação, e podem ser visualizados utilizando o *Genome browser* integrado (MCCLURE *et al.*, 2012).

5. Resultados e Discussão

5.1 Análise de qualidade

Para cada posição um gráfico tipo ‘BoxWhisker’ é desenhado. Os elementos da trama são como se segue: a linha vermelha central é o valor da mediana, a caixa amarela representa o intervalo inter-quartil (25-75%), os traços superiores e inferiores representam os pontos de 10% e 90%, a linha azul representa a qualidade média.

O gráfico seguinte ilustra uma análise de qualidade dos dados brutos de uma corrida de RNA-Seq da condição Nitrato. Pode-se observar que a qualidade das bases diminui quando se aproxima do fim.

O eixo y no gráfico mostra os índices de qualidade. Quanto maior a pontuação, melhor a qualidade da base. O pano de fundo do gráfico divide o eixo y em camadas de muito boa qualidade (verde), camadas de qualidade razoável (laranja) e camadas de má qualidade (vermelho). A qualidade das bases, na maioria dos sequenciadores, irá diminuir enquanto a corrida progride, por isso é comum ver que a qualidade das bases caem para a área em laranja no final de uma leitura.

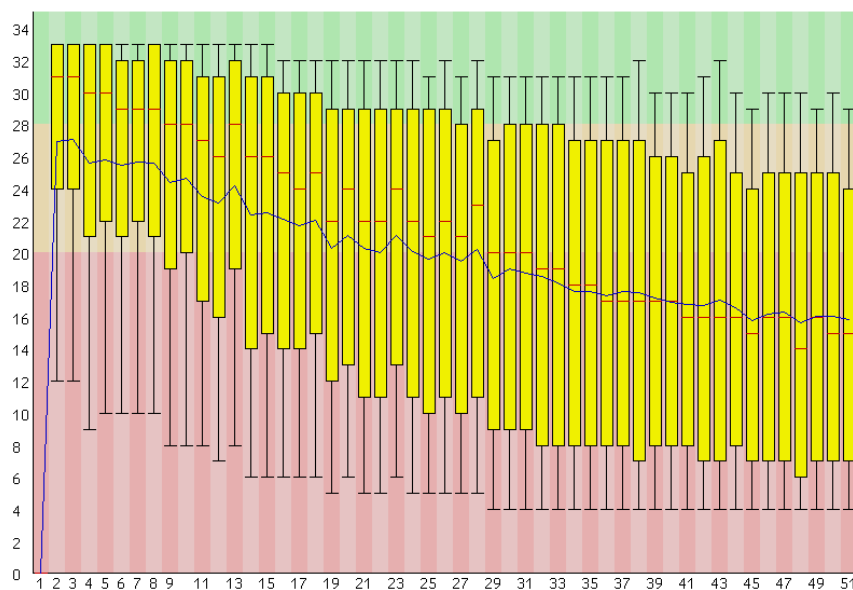


FIGURA 05. Gráfico tipo BoxWhisker dos dados ‘brutos’ do sequenciamento de uma amostra da condição Nitrato. O eixo y mostra os índices de qualidade. O eixo x representa cada base.

Após aplicar o filtro com a média de qualidade das bases maior que 20 e excluir as 10 últimas bases (*triming*), o gráfico de análise de qualidade passa a ser o apresentado na Figura 06.

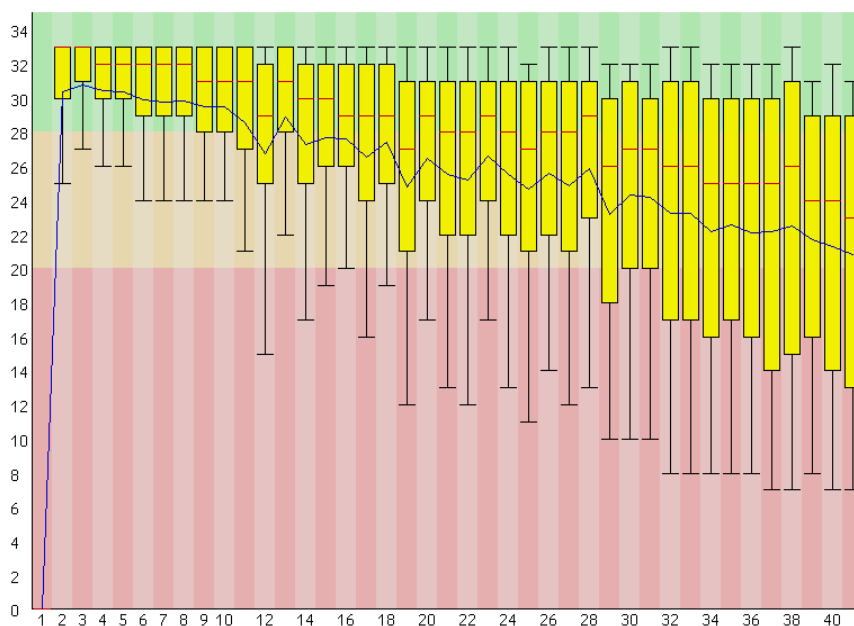


FIGURA 06. Gráfico tipo BoxWhisker dos dados após trinar as últimas 10 bases e realizar o filtro de qualidade Q20 do sequenciamento de uma amostra da condição Nitrato. O eixo y mostra os índices de qualidade. O eixo x representa cada base.

5.2 Identificação dos genes essenciais de *H. seropedicae* utilizando os dados de transcriptoma de onze condições de cultivo diferentes.

Os genes conhecidos como *housekeeping* estão envolvidos na manutenção das funções celulares básicas, seus produtos são essenciais para que uma célula possa funcionar. É esperado que esses genes apresentem um nível de expressão constante em todas as células e em todas as condições (EISENBERG & LEVANON, 2003; DU *et al.*, 2013), embora possa existir uma variação na expressão em determinadas condições. A identificação desses genes permite um melhor entendimento do funcionamento básico da bactéria (GREER *et al.*, 2010).

Primeiramente foram feitos vários testes para estabelecer quais os parâmetros que seriam utilizados para criar o conjunto de genes que estivessem sempre expressos nas 11 condições testadas. Na figura 07 está mostrado o fluxograma que foi seguido para a criação desse conjunto de genes. Neste trabalho, foi realizado o aparaamento das sequências para o tamanho de 40 bases utilizando *script* em Perl (Anexos 5 e 6) e o filtro de qualidade Q20.

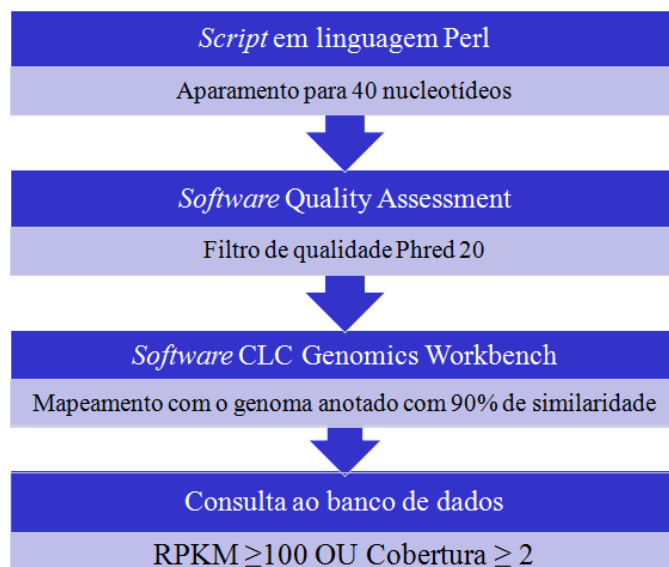


FIGURA 07. Fluxo de tratamento dos dados de RNA-Seq.

Os transcritos das duas replicatas de cada condição foram mapeados em relação ao genoma anotado (arquivo.gbk) simultaneamente, usando o programa CLC Workbench. Os genes de rRNA e tRNA foram excluídos. Os parâmetros utilizados para mapeamento estão mostrados na tabela 02 e os números de *reads* mapeados em cada condição estão mostrados na tabela 03. Cada condição de RNA-Seq possui uma biblioteca de tamanho diferente, devido aos níveis de expressão gênica necessários para cada condição ser diferente, também por possíveis vieses na própria técnica de sequenciamento e de análise dos dados (OSHLACK & WAKEFIELD, 2009).

TABELA 02. Parâmetros para mapeamento dos *reads* com o CLC Workbench.

Custo de Não-alinhamento	2
Custo de Inserção	2
Custo de Deleção	2
Fração de Tamanho	0,9
Fração de Similaridade	0,9
Custo erro de cor	3
Manuseio de Alinhamento não específico	Ignorar

TABELA 03. Tamanhos das bibliotecas de cada condição de cultivo e a quantidade de *reads* mapeados.

Condições	Tamanho das Bibliotecas	Reads mapeados
NFbHPN-Malato	38.086.709	1.226.182
NFbHPN-Malato-Naringenina	18.475.948	832.665
NFbHPN-Malato-Alto Oxigênio	4.011.647	1.324.500
NFbHPN-Malato-Baixo Oxigênio	6.843.252	1.316.022
NFbHPN-Malato-Nitrato	54.140.643	4.629.469
Planctônico–milho-1dia	19.171.260	725.832
Planctônico-milho-3dias	18.360.305	916.216
Aderido–milho- 1dia	27.561.077	273.893
Aderido-milho3dias	14.491.361	569.907
Planctônico-trigo	55.989.233	1.685.612
Aderido-trigo	15.573.219	116.373

O mapeamento foi exportado do CLC Workbench em forma de tabela e importado para um banco de dados e então foram selecionados os genes que estavam expressos nas 11 condições com $RPKM \geq 100$ OU Cobertura ≥ 2 . Foi retornado um conjunto de 583 genes, que listado no Anexo 1. Foram considerados o RPKM e a Cobertura como parâmetros para corte. Esses parâmetros de corte foram determinados por serem os valores mínimos para um gene ser confiavelmente considerado expresso, excluindo assim os ruídos transcricionais (BULLARD *et al.*, 2010 ; DIJK *et al.*, 2012 ; CROUCHER & THOMPSON, 2010). Levando também em consideração que os dados já passaram por um filtro de qualidade.

A tabela 04 traz a relação dos 50 genes mais expressos nas onze condições de transcriptoma, com seus respectivos produtos e vias metabólicas das quais participam ou função de acordo com a classificação COG. Todos os genes presentes na tabela possuem $RPKM \geq 590$ em todas as condições.

Ao observar o papel desses genes podemos perceber que a maioria realiza funções essenciais para a manutenção da vida da bactéria. Vale resaltar que as informações contidas no KEGG estão desatualizadas, podendo algum gene ter função conhecida ou uma segunda função.

TABELA 04. Relação dos 50 genes mais expressos nas onze condições de transcriptoma, com seus respectivos produtos e vias metabólicas das quais participam ou função. Todos os genes presentes na tabela possuem RPKM ≥ 590 em todas as condições.

Nº no Genoma	Produto do Gene	Via(s) metabólica(s) que participa e função(ões)
Hsero_0068	Hsero_0068 – proteína hipotética conservada.	Via metabólica não encontrada pelo KEGG e função desconhecida.
Hsero_0097	tufB - EF-Tu - fator de alongamento de proteína.	Via metabólica não encontrada pelo KEGG e função de tradução, estrutura ribossomal e biogênese.
Hsero_0099	secE - pré-proteína translocase secE proteína de subunidade transmembranar.	Exportação de proteína. Sistema de secreção bacteriana.
Hsero_0103	rplJ - 50s ribossomal proteína subunidade L10.	Proteína ribossomal.
Hsero_0104	rplL - 50s ribossomal proteína subunidade L7/L12.	Proteína ribossomal.
Hsero_0111	fusA – fator de alongamento de proteína G1 (EF-G1).	Via metabólica não encontrada pelo KEGG e função de tradução, estrutura ribossomal e biogênese.
Hsero_0326	Hsero_0326 - Proteína de ligação ao DNA da família histona HU.	Via metabólica não encontrada pelo KEGG e função de replicação, recombinação e reparação do DNA.
Hsero_0364	rplU - 50s ribossomal proteína subunidade L21.	Proteína ribossomal.
Hsero_0381	nrdB - proteína transmembrana ribonucleosídeo redutase (cadeia pequena)	Metabolismo de pirimidina. Metabolismo de purina.
Hsero_0505	cspC – choque a frio proteína reguladora de transcrição.	Via metabólica não encontrada pelo KEGG. Função de transcrição.
Hsero_0870	rpoD – RNA Polimerase proteína reguladora fator de transcrição sigma-70	Via metabólica não encontrada pelo KEGG. Função de transcrição.
Hsero_0997	groES1 - GroES (HSP10)	Via metabólica não encontrada pelo KEGG.

Nº no Genoma	Produto do Gene	Via(s) metabólica(s) que participa e função(ões)
	de proteína de co-chaperonina	Função de Modificação pós-traducional, turnover de proteínas, chaperonas.
Hsero_1181	Hsero_1181 - estearoil-CoA desaturase proteína oxidoredutase.	Metabolismo de ácidos graxos. Biossíntese de ácidos gordos insaturados.
Hsero_1383	hflB - ATP-dependente metalopeptidase Zn (divisão celular FtsH) proteína transmembrana.	Via metabólica não encontrada pelo KEGG. Função de Modificação pós-traducional, turnover de proteínas, chaperonas.
Hsero_1393	rpsT - subunidade protéica ribossomal 30S.	Proteína ribossomal.
Hsero_1397	cspD - proteína domínio de ligação do DNA a choque a frio.	Via metabólica não encontrada pelo KEGG. Função de transcrição.
Hsero_1398	clpS - proteína adaptadora de protease Clp ATP-dependente .	Via metabólica não encontrada pelo KEGG. Função de Modificação pós-traducional, turnover de proteínas, chaperonas.
Hsero_1399	clpA - proteína Protease dependente de ATP (ATP-ligação subunidade de especificidade)	Via metabólica não encontrada pelo KEGG. Função de Modificação pós-traducional, turnover de proteínas, chaperonas.
Hsero_1639	Hsero_1639 - proteína família Phasin.	Via metabólica não encontrada pelo KEGG. Função desconhecida.
Hsero_1921	rpmF – proteína ribossomal subunidade L32 50S	Proteína ribossomal. Função de tradução, estrutura ribossomal e biogênese.
Hsero_1926	acpP - proteína transportadora de acil.	Via metabólica não encontrada pelo KEGG. Função de metabolismo e transporte de lipídios. Biossíntese de metabolitos secundários, transporte e catabolismo.
Hsero_1929	rpoE - RNA polimerase fator sigma-E (sigma-24) proteína reguladora da transcrição	Via metabólica não encontrada pelo KEGG. Função de transcrição.
Hsero_1978	infC - fator de inicio da tradução de proteínas	Via metabólica não encontrada pelo KEGG. Função de tradução, estrutura ribossomal e biogênese.
Hsero_2062	Hsero_2062 – proteína YcgR e PilZ famílias (tipo IV montagem pilus PilZ)	Via metabólica não encontrada pelo KEGG. Função de mobilidade celular, tráfego intracelular, secreção e transporte vesicular.
Hsero_2069	flaG - proteína flagelina	Via metabólica não encontrada pelo KEGG. Função de mobilidade celular.
Hsero_2070	fliC – proteína flagelina, proteína estrutural do filamento.	Sistema de 2 componentes. Montagem Flagelar.

Nº no Genoma	Produto do Gene	Via(s) metabólica(s) que participa e função(ões)
Hsero_2071	Hsero_2071 – proteína hipotética.	Via metabólica não encontrada pelo KEGG. Função desconhecida.
Hsero_2178	rpsB – proteína S2 subunidade ribossomal 30S.	Proteína ribossomal.
Hsero_2580	lon - Protease ATP-dependente de proteína LA	Via metabólica não encontrada pelo KEGG. Função de Modificação pós-traducional, turnover de proteínas, chaperonas.
Hsero_2905	Hsero_2905 – proteína hipotética conservada.	Via metabólica não encontrada pelo KEGG. Função desconhecida.
Hsero_2948	hfq - Proteína RNA chaperona Hfq.	Degradação de RNA.
Hsero_2956	ndk - proteína quinase nucleósido-difosfato	Metabolismo de pirimidina. Metabolismo de purina. Biossíntese de metabólitos secundários.
Hsero_2981	cheW - Chew proteína quimiotaxia, regulador positivo da atividade da proteína Chea.	Sistema de 2 componentes. Quimiotaxia bacteriana.
Hsero_2988	flhC - transcrição flagelar proteína reguladora ativador.	Sistema de 2 componentes. Montagem Flagelar.
Hsero_2989	flhD - transcrição flagelar proteína reguladora ativador.	Sistema de 2 componentes. Montagem Flagelar.
Hsero_3028	cspD - proteína reguladora de transcrição tipo choque a frio	Via metabólica não encontrada pelo KEGG. Função de transcrição.
Hsero_3063	efp - tradução alongamento fator P (EF-P) / tradução iniciação fator 5A (FEI-5A)	Via metabólica não encontrada pelo KEGG. Função de tradução, estrutura ribossomal e biogênese.
Hsero_3196	Hsero_3196 – proteína hipotética.	Via metabólica não encontrada pelo KEGG. Função desconhecida.
Hsero_3499	qor - NADPH: quinona oxidoredutase proteína	Via metabólica não encontrada pelo KEGG. Função de produção e conversão de energia.
Hsero_3500	Hsero_3500 – proteína hipotética.	Via metabólica não encontrada pelo KEGG. Função desconhecida.
Hsero_3688	ihfB - subunidade do fator de integração hospedeiro (IHF-beta) proteína	Via metabólica não encontrada pelo KEGG. Replicação de DNA, recombinação e reparação.
Hsero_3696	Hsero_3696 – proteína péptido sinal.	Via metabólica não encontrada pelo KEGG. Função parede celular.
Hsero_3845	Hsero_3845 – proteína hipotética conservada.	Via metabólica não encontrada pelo KEGG. Função desconhecida.

Nº no Genoma	Produto do Gene	Via(s) metabólica(s) que participa e função(ões)
Hsero_3850	rplM – proteína ribossomal subunidade L13 50S	Proteína ribossomal.
Hsero_3857	Hsero_3857 - sistema de transporte tipo ABC de glutamato / aspartato, proteína componente periplásmicos	Sistema de 2 componentes. Transportadores ABC.
Hsero_3974	yhbH - proteína modulação Sigma-54	Via metabólica não encontrada pelo KEGG. Função de tradução, estrutura ribossomal e biogênese.
Hsero_4198	Hsero_4198 - NAD_dependente epimerase / desidratase proteína	Via metabólica não encontrada pelo KEGG. Função parede celular.
Hsero_4295	Hsero_4295 - membrana externa (porina) proteína	Via metabólica não encontrada pelo KEGG. Função parede celular.
Hsero_4367	atpE - Tipo F0F1 ATP sintase, subunidade c proteína	Fosforilação oxidativa.
Hsero_4678	Hsero_4678 – proteína hipotética desconhecida.	Via metabólica não encontrada pelo KEGG. Função desconhecida.

5.3 Classificação Funcional das proteínas codificadas pelos genes identificados

Utilizando os recursos do COG (The database of Clusters of Orthologous Groups of proteins), foi realizada a classificação funcional das 583 proteínas que estão expressas nas onze condições já descritas, de acordo com os critérios estabelecidos. O resultado da classificação do COG pode ser visualizado na figura 08:

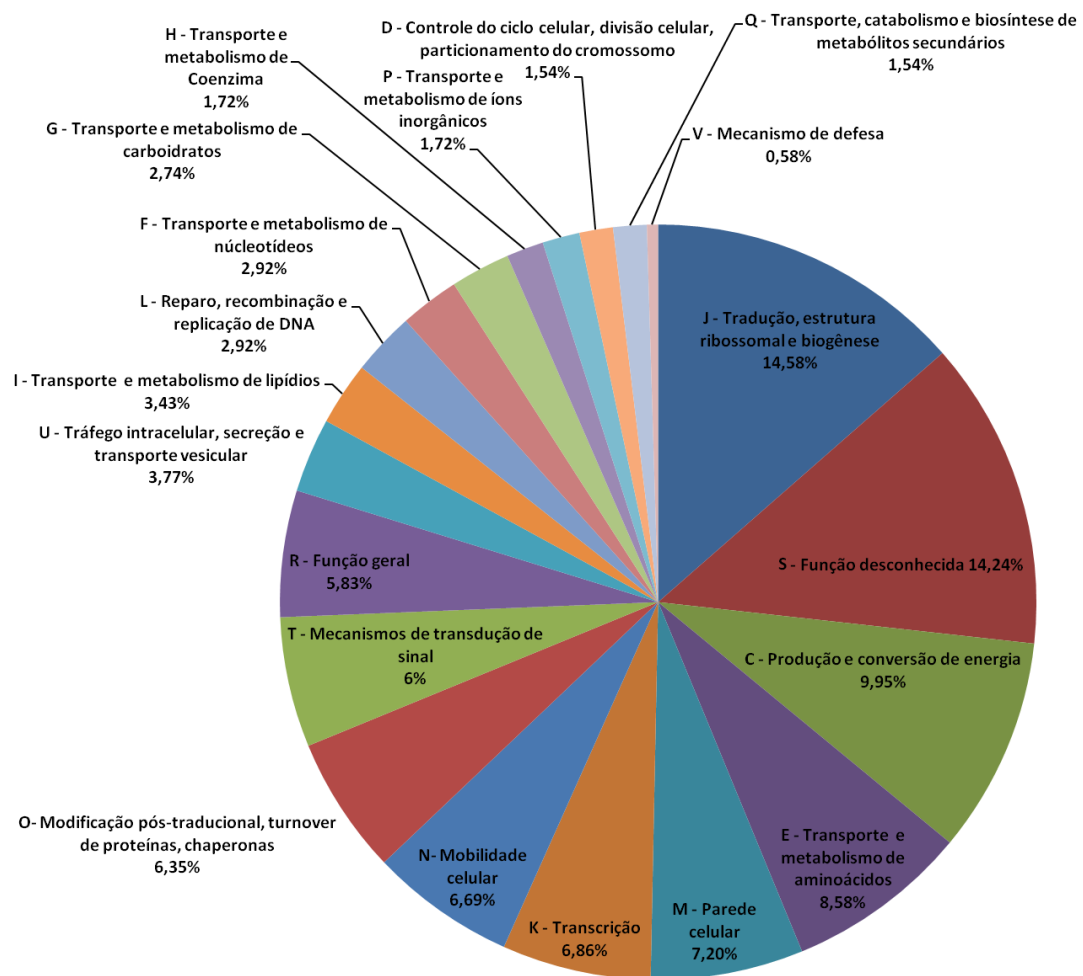


FIGURA 08. Classificação funcional realizada pelo COG dos 583 genes mais expressos em todas as condições analisadas.

Todas as categorias funcionais do COG estão cobertas pelos 583 genes identificados. A maioria (14,58%) das proteínas expressas faz parte da categoria funcional tradução, estrutura e biogênese ribossomal, essas vias são essenciais para a manutenção celular. 14,24% das proteínas não tem função conhecida, isso indica que proteínas que são essenciais para a célula ainda não tiveram a sua função determinada.

Também foi realizada a classificação funcional das proteínas codificadas pelos 50 genes mais expressos em todas as condições (Figura 09). Os resultados mostram que as duas classes funcionais mais representadas são as mesmas obtidas quando são utilizadas as 583 proteínas.

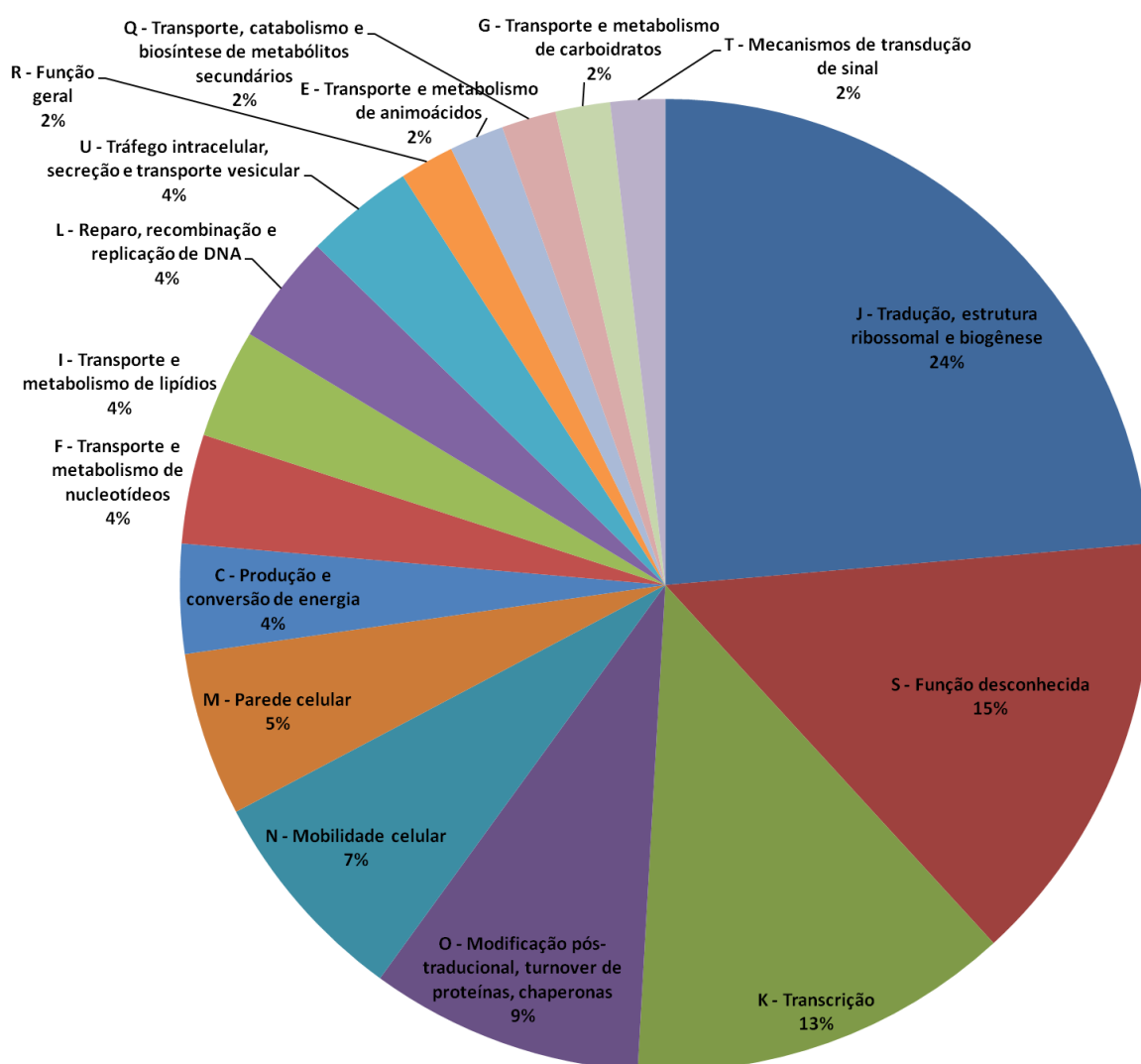


Figura 09. Classificação funcional realizada pelo COG dos 50 genes mais expressos em todas as condições analisadas.

5.4 Distribuição das proteínas codificadas pelos genes identificados nas principais vias metabólicas

Tendo em mãos o conjunto de genes expressos nas 11 condições de transcriptoma de acordo com o critério já exposto, procuramos identificar quais vias metabólicas estão ativas nas 11 condições. Foram utilizados os recursos do KEGG para esta atividade. Nós identificamos 7 vias metabólicas que parecem estar ativas em todas as condições testadas: ciclo do Ácido Cítrico, metabolismo do Piruvato, degradação de RNA, vias de montagem do Flagelo, Glicólise, Gliconeogênese e Fosforilação Oxidativa. E também as proteínas ribossomais.

As vias metabólicas do ciclo do ácido cítrico (ciclo de Krebs), metabolismo do piruvato e fosforilação oxidativa são essenciais para a manutenção da vida, pois fazem parte do processo de produção de energia.

A glicose é uma fonte de energia para o microorganismo. Após atravessar a membrana do microorganismo a glicose é fosforilada a glicose-6-fosfato, que através da glicólise levará à produção de piruvato, que pode ser convertido a Acetil CoA e utilizado pelo ciclo de Krebs.

As vias de montagem flagelar são utilizadas para a locomoção da bactéria sendo assim essenciais (ALDRIDGE & HUGHES, 2002 ; AIZAWA, 2009). A degradação de moléculas de RNA também é um mecanismo comum para a célula bacteriana (HOUSELEY & TOLLERVEY, 2009).

A figura 11 mostra a via metabólica do ciclo do Ácido Cítrico ou ciclo de Krebs, com os genes constitutivos em destaque (vermelho). Apenas o gene da etapa 4.2.1.2 estava com RPKM abaixo de 100 e cobertura menor que 2 na condição Aderido–milho- 1dia dos transcriptomas. Por considerar que o nível de expressão é satisfatório, ele foi incluso no conjunto. O mesmo foi feito para os genes de montagem flagelar, glicólise e fosforilação oxidativa. Ou seja, os genes que estavam próximos da linha de corte, cuja expressão estava diferente de zero e são muito importantes para as funções vitais do organismo, foram incluídos.

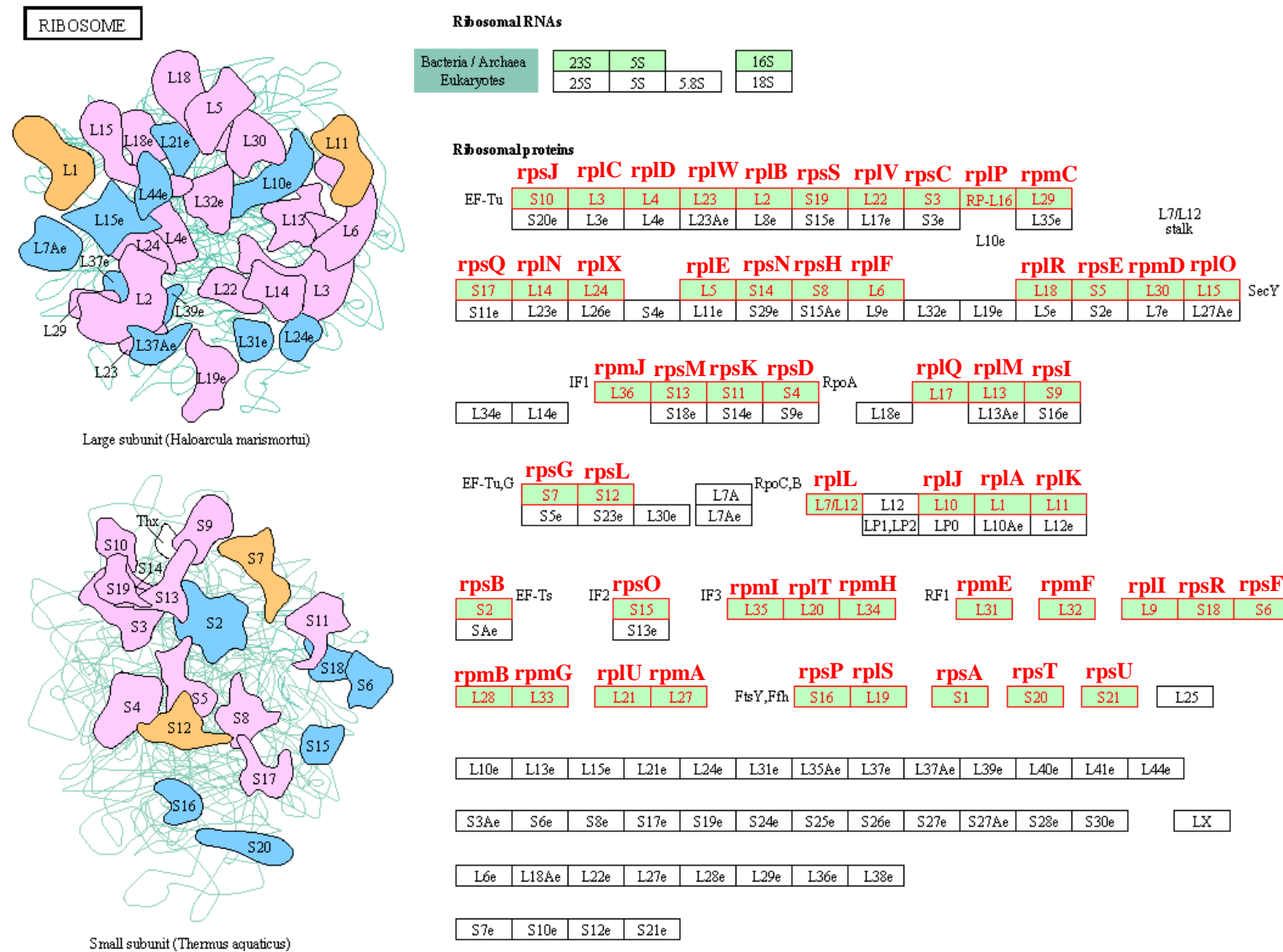


FIGURA 10. Proteínas ribossomais de *H. seropedicae*. Todas estão presentes entre os genes mais expressos, representados em caixa verde com letras vermelhas. FONTE: adaptado de KEGG Pathway.

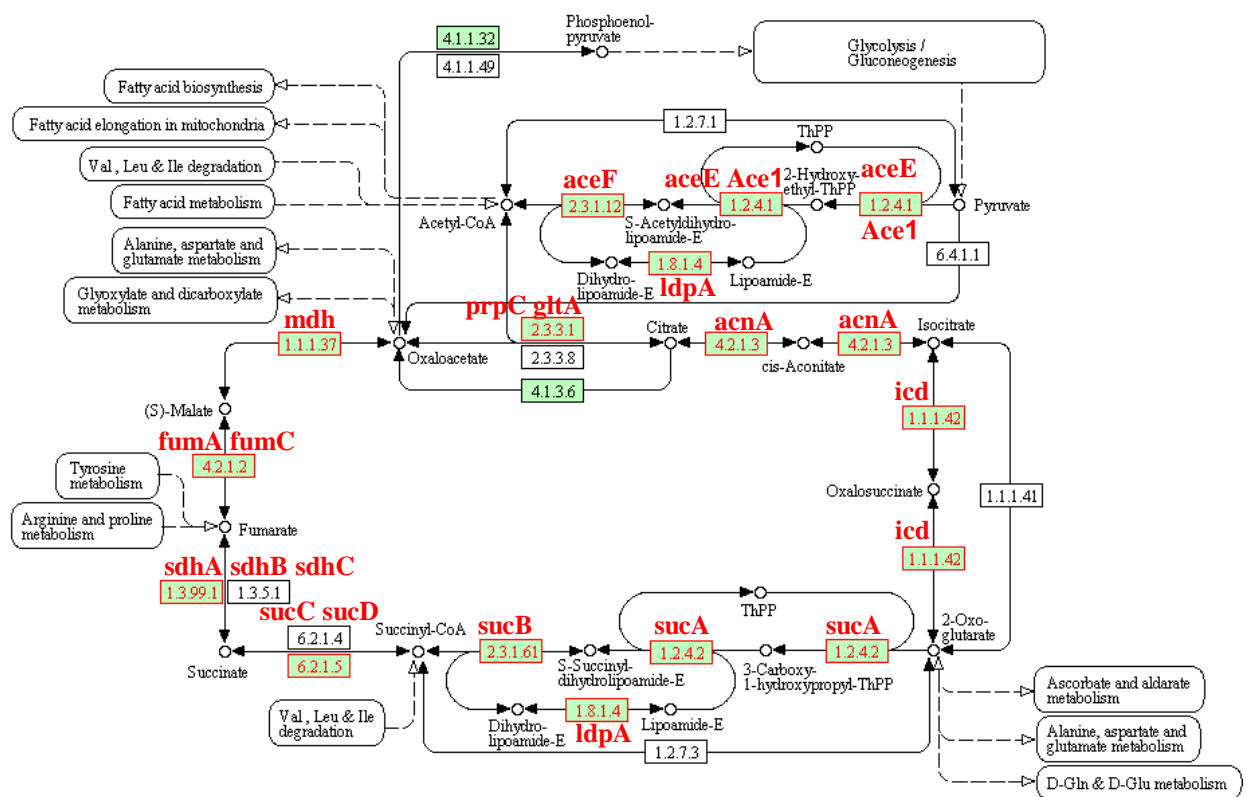
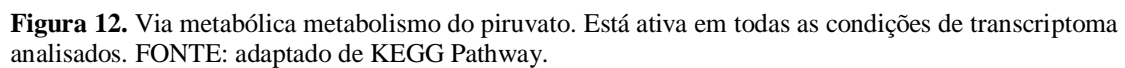


FIGURA 11. Via metabólica ciclo de Krebs. Esta via está ativa em todas as condições de transcriptoma analisados. O conjunto de genes essenciais estão representados em caixa verde com letras vermelhas e os genes que estão presentes no genoma de *H. seropedicae* em caixa verde com letras pretas. FONTE: adaptado de KEGG Pathway.



The diagram illustrates the components and organization of RNA degradation pathways. At the top, a linear schematic shows the **RNase E (Escherichia coli)** structure, divided into a **Catalytic domain** and a **C-terminal domain** (labeled as **Degradosome component interaction**). The catalytic domain contains the **RNA helicase (RhIb)** and **Enolase** subunits, while the C-terminal domain contains the **PNPase** subunit. Below this, a 3D model of the **RNA degradosome** is shown, with RNA (5' end) being degraded. The **pyrophosphatase RppH** is indicated as acting on the 5' end of the RNA. To the right, four types of RNA degradosomes are listed with their associated proteins in color-coded boxes:

- RNA degradosome type A (Escherichia coli):** RNase E (red), RhIb (white), Enolase (red), PNPase (red).
- RNA degradosome type B (Pseudomonas):** RNase E (red), RhIe (white), RNase R (green).
- RNA degradosome type C (Rhodobacter):** RNase E (red), helicases (white), Rho (red).
- RNA degradosome type D (Bacillus subtilis):** PNPase (red), RNase R (green), RNase J (white).

At the bottom, **Associated proteins** are listed in color-coded boxes: DnaK (red), GroEL (green), Hfq (red), PPK (green), and PAP (green).

FIGURA 13. Via metabólica degradação de RNA. FONTE: adaptado de KEGG Pathway.

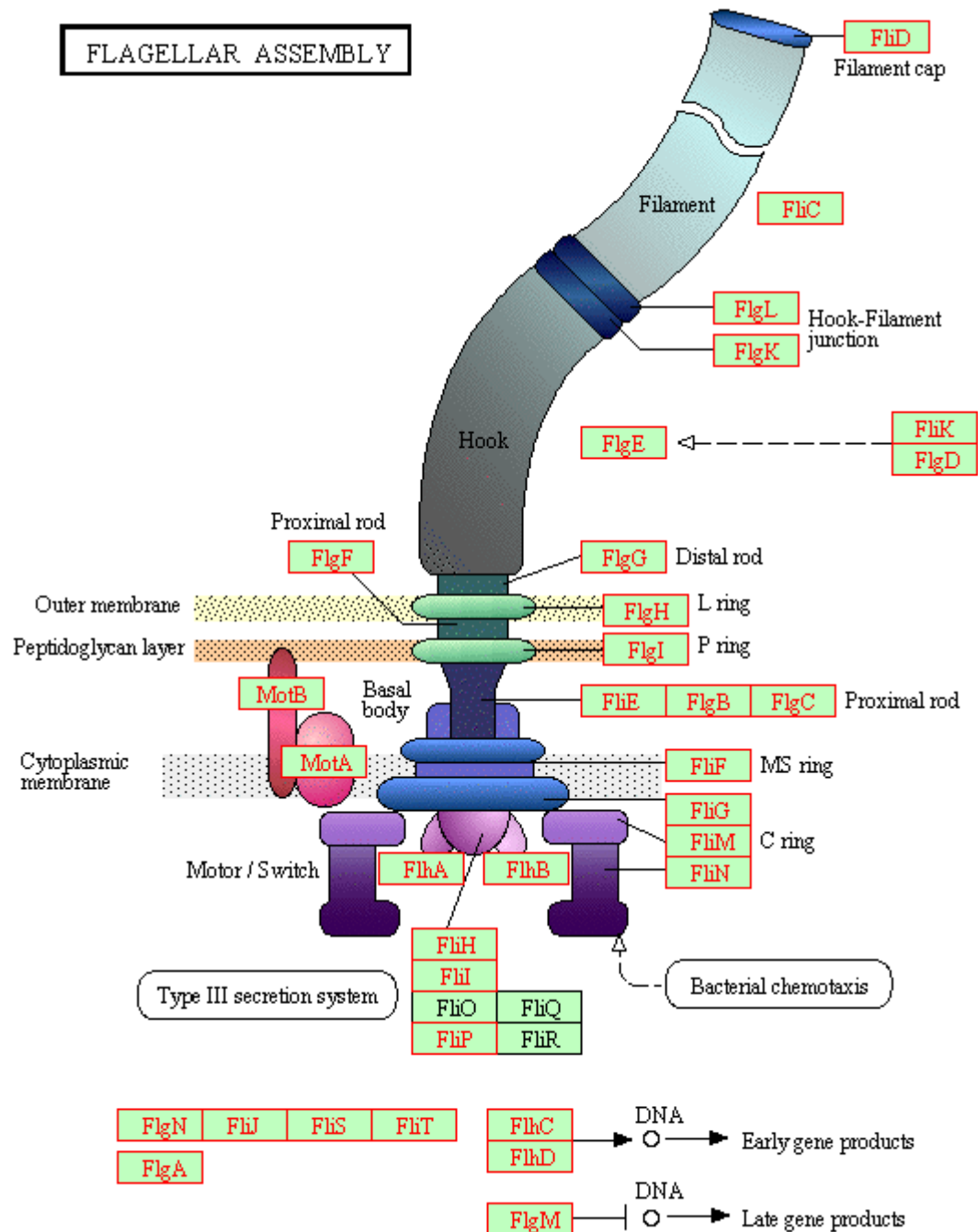


FIGURA 14. Via metabólica montagem flagelar. FONTE: adaptado de KEGG Pathway.

Na montagem flagelar (Figura 14) os gene **FliO** e **FliQ** possuem RPKM e Cobertura iguais a zero na condição Raiz de Trigo. O gene **FliR** possui RPKM e Cobertura iguais a zero na condição Raiz de Trigo e Aderido 1 dia a raiz do milho, portanto não foram adicionados ao conjunto de genes constitutivos.

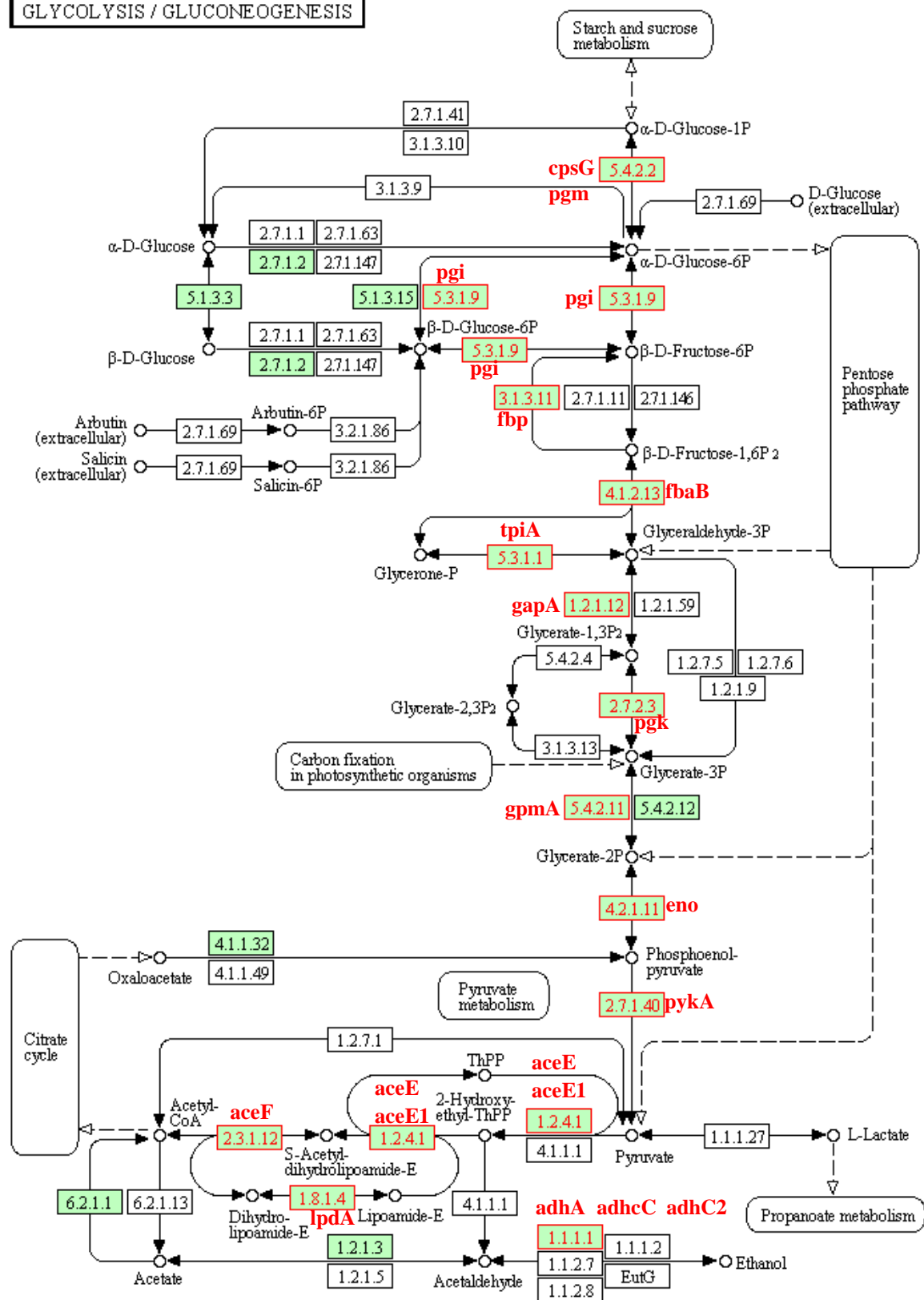


FIGURA 15. Via metabólica da glicose. Esta via está ativa em todas as condições de transcrito analisados. FONTE: adaptado de KEGG Pathway.

OXIDATIVE PHOSPHORYLATION

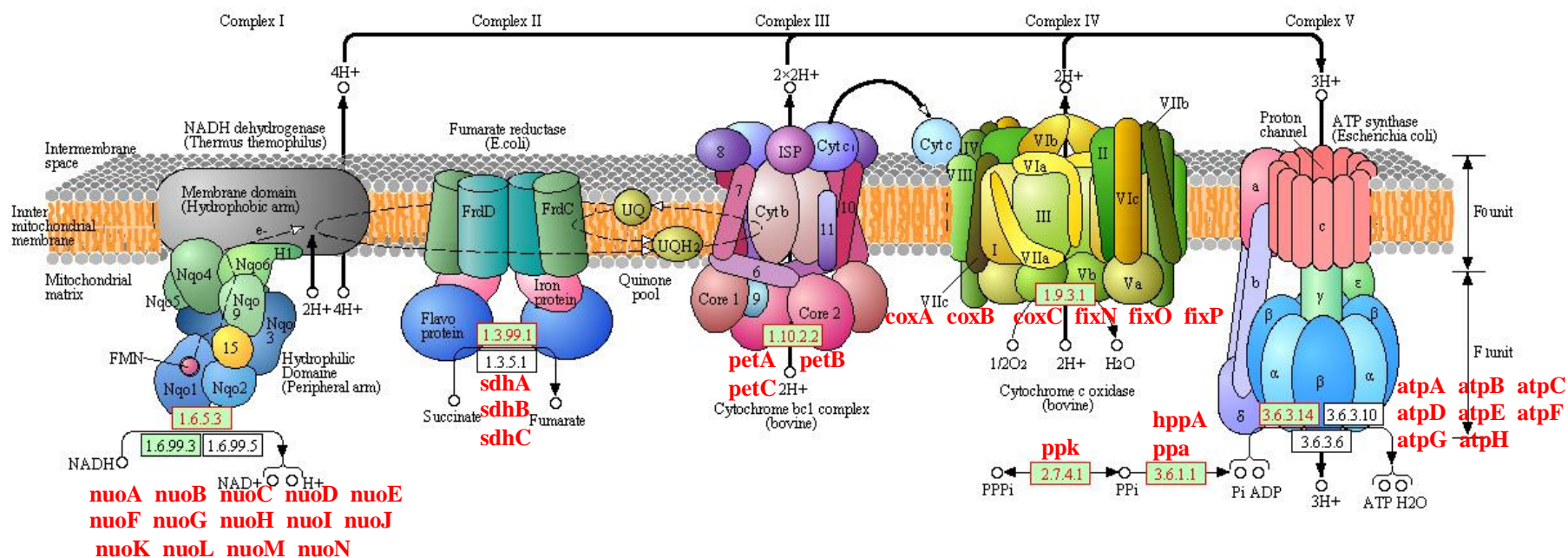


FIGURA 16. Via metabólica fosforilação oxidativa. Esta via está ativa em todas as condições de transcrito analisados. FONTE: adaptado de KEGG Pathway.

5.5 Análise dos 538 genes identificados com o *software* CodonW

Buscando saber se existe uma preferência de uso de códons entre o conjunto de genes sempre expressos (583) e os demais (4157) foram realizadas análises estatísticas. Foram excluídos os rRNAs e os tRNAs.

Os índices CAI (índice de adaptação de códon), CBI (índice de tendência de códon), Fop (frequência de códons ótimos) e GC3 (Guanina e Citosina na terceira posição do códon) foram analisados pelo teste *t* de Student e expressos como média±erro padrão da média.

A tabela 05 mostra a média e o erro padrão da média dos índices CAI, CBI, Fop e GC3s respectivamente para o conjunto de genes sempre expressos. E a tabela 06 mostra a média e o erro padrão da média dos índices CAI, CBI, Fop e GC3s para os demais genes.

TABELA 05. Média e erro padrão da média dos índices CAI, CBI, Fop e GC3 dos genes sempre expressos.

	CAI	CBI	Fop	GC3s
Média	0,3838	0,3159	0,6015	0,7775
Erro Padrão da Média	0,0033	0,0054	0,0031	0,0042

TABELA 06. Média e erro padrão da média dos índices CAI, CBI, Fop e GC3 dos demais genes.

	CAI	CBI	Fop	GC3s
Média	0,3313	0,2896	0,5806	0,8117
Erro Padrão da Média	0,0008	0,0015	0,0008	0,0013

Com os dados sobre média e erro padrão da média para os índices dos genes sempre expressos e os demais foram feitos gráficos de colunas com o resultado do teste *t* de Student.

O teste *t* de Student mostrou que a diferença dos valores de média e erro padrão da média entre os dois conjuntos é estatisticamente significativa com valor de $P < 0,0001$ (*).

A figura 17 mostra os gráficos de colunas resultantes do teste *t* de Student para o índice a) CAI, b) CBI, c) Fop e d) GC3.

Esse resultado indica haver uma preferência de códon entre os genes sempre expressos e os demais genes.

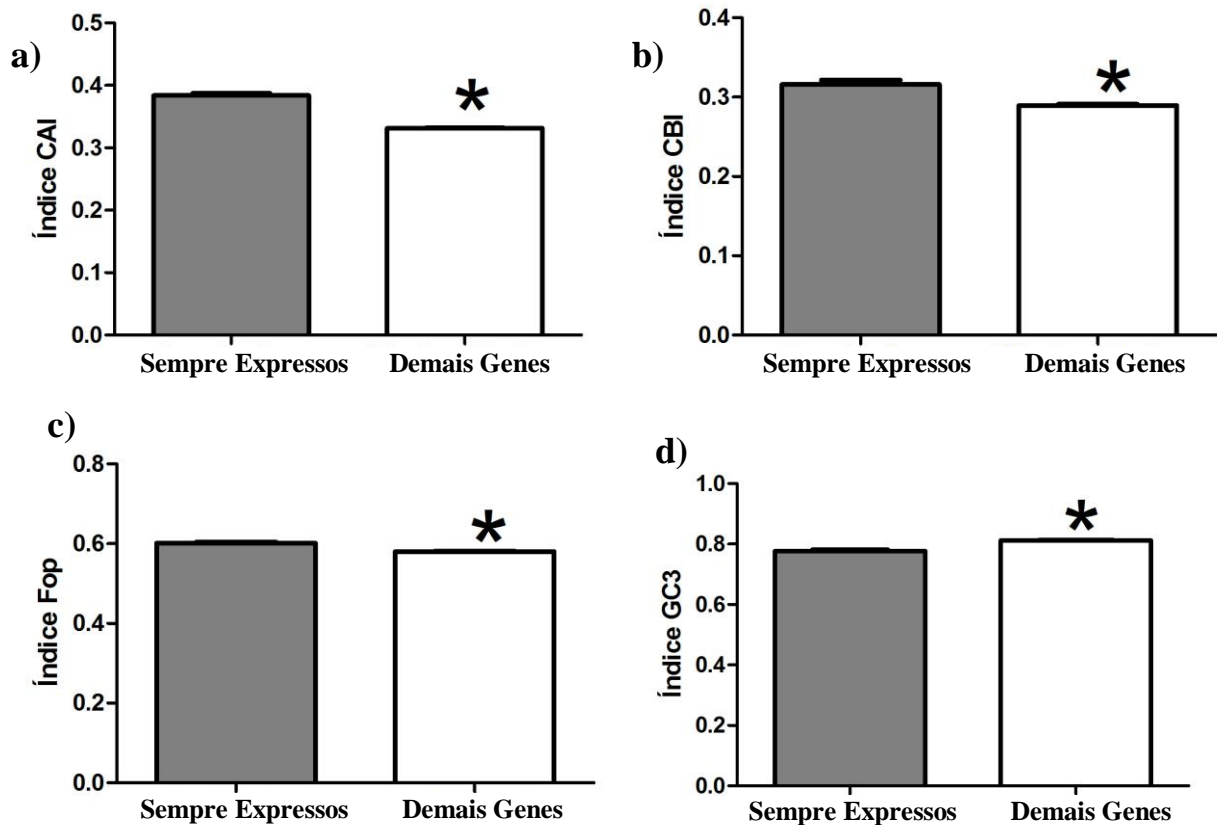


FIGURA 17. Gráfico de colunas dos índices CAI (a), CBI (b), Fop (c) e GC3 (d). Houve diferenças estatisticamente significantes entre os genes sempre expressos e os demais genes em todos os índices com $P < 0,0001$. Dados analisados pelo Teste *t* de Student e expressos como média±erro padrão da média.

5.6 Identificação dos operons presentes em *H. seropedicae* utilizando dados de transcriptoma

Operons são formados por genes próximos que estão sendo transcritos a partir do mesmo promotor. Os dados de RNA-Seq podem ser utilizados para definir quais genes podem formar um operon e utilizando esta estratégia foram preditos 360 operons em *Mycobacterium marinum* (WANG *et al.*, 2009).

Para identificar os operons de *H. seropedicae* foi utilizado o *software* Rockhopper, que fez um levantamento de 1423 prováveis operons, mas confirmou 861 ao analisar os dados de RNA-Seq. Foram passados para o programa os arquivos de RNA-Seq (csfasta ou fastq) de cada condição separadamente com replicatas e não foram observadas diferenças na predição de operons entre as 11 condições de transcriptoma e nem entre os dados brutos e os dados com filtro de qualidade Q20 e trimados.

Este resultado pode não representar cem por cento dos operons de *H. seropedicae*, uma vez que alguns operons preditos por outros softwares com base em outros métodos não aparecem como confirmados nessa análise (TABOADA *et al.*, 2012 ; SOLOVYEV & SALAMOV, 2011 ; MAO, 2013) , pois o Rockhopper não explora a possibilidade de genes a serem co-transcritos em algumas condições e outras não. Também não investiga operons controlados por vários promotores (MCCLURE *et al.*, 2013), possuindo uma sensibilidade que varia de 89-96% e especificidade variando de 86-95%.

A figura 18 mostra uma visualização no Artemis da expressão de um operon de quatro genes (destacados pela cor rosa) *nuoA*, *nuoB*, *nuoC* e *nuoD*. Os dois genes seguintes que aparecem na imagem (*nuoE* e *nuoF*) são expressos juntos em outro operon. Esses genes fazem parte da via metabólica da fosforilação oxidativa.

De acordo com o banco de dados de operons procariotos (DOOR – Database of prokaryotic Operons) 193 espécies possuem esses genes em forma de operon, mas em alguns organismo (como a *Xylella fastidiosa Temecula1*) o operon é composto de quatorze genes (*nuoA* – *nuoN*). Em outros organismos (como *Burkholderia pseudomallei* 668) apenas os quatros genes são expressos juntos.

O DOOR estabelece operons para 2072 genomas bacterianos com base em inferências computacionais com precisão de aproximadamente 90% avaliado por BROUWER *et al.* (2008).

Em *H. seropedicae* SMR1, para a sequência de genes *nuoA* – *nuoN* acontece os operons conforme a tabela 08.

TABELA 07. Operons para os genes *nuoA-nuoN* preditos pelo programa Rockhopper.

Início da transcrição	Fim da transcrição	Fita do DNA	Nº de genes	Nome dos genes
1998086	2000784	+	4	<i>nuoA</i> , <i>nuoB</i> , <i>nuoC</i> , <i>nuoD</i>
2000892	2002663	+	2	<i>nuoE</i> , <i>nuoF</i>
2002746	2006647	+	3	<i>nuoG</i> , <i>nuoH</i> , <i>nuoI</i>
2006788	2007774	+	2	<i>nuoJ</i> , <i>nuoK</i>
2007890	2011493	+	2	<i>nuoL</i> , <i>nuoM</i>
2011652	2013487	+	2	<i>nuoN</i> , <i>Hsero_1781</i>

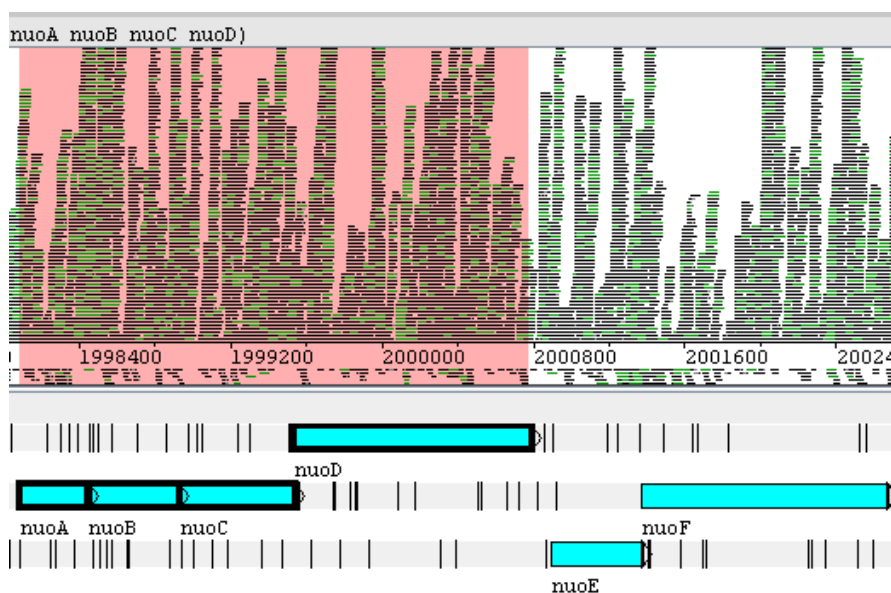


FIGURA 18. Visualização no Artemis da expressão do operons *nuoA-nuoD* na condição NfbHPN-Malato-Naringenina.

5.7 Identificação de genes de *H. seropedicae* ainda não anotados utilizando dados de RNAseq

A análise dos dados de RNA-Seq mostrou que havia transcritos em regiões não anotadas no sequenciamento e anotação do genoma. Alguns desses transcritos estão em regiões de ORFs (Open Reading Frame – Fase Aberta de Leitura) e outros não. Então a anotação desses transcritos como genes codificadores de proteína ou como ncRNAs (LEHNINGER *et al.*, 2008).

5.7.1 HGF

Foi realizada nova anotação, a partir do genoma no formato fasta, utilizando o programa HGF (Hybrid Gene Finder – Localizador Híbrido de Gene). Disponível no sítio (<http://www.bioinfo.ufpr.br/hgf/>), (RAITZ, 2010). Foram preditos 422 genes que ainda não havia sido anotados no genoma de *H. seropedicae*, sendo que 275 ainda não estão anotados em nenhum organismo no NCBI. Essa anotação foi realizada em novembro de 2013 e o programa gerou um arquivo com anotação dos genes, os que já previamente anotados no genoma de *H. seropedicae* foram excluídos usando um *script* em MatLab.

As 422 proteínas hipotéticas encontradas pelo *software* HGF foram mapeadas nas 11 (onze) condições de transcriptoma, sendo antes aparados em dez nucleotídeos no sentido 3' – 5' e o filtro de qualidade Q20. Foram detectados 340 genes expressos com cobertura maior ou igual a três. A tabela 08 mostra a quantidade de novos genes hipotéticos confirmados em cada condição de transcriptoma.

TABELA 08. Quantidade de novos genes hipotéticos por condição.

Condições	Cobertura ≥ 3
NFbHPN-Malato	248
NFbHPN-Malato-Naringenina	238
NFbHPN-Malato-Alto Oxigênio	188
NFbHPN-Malato-Baixo Oxigênio	218
NFbHPN-Malato-Nitrato	287
Planctônico–milho-1dia	139
Planctônico-milho-3dias	156
Aderido–milho- 1dia	88
Aderido-milho3dias	123
Planctônico- trigo	214
Aderido-trigo	54

No Anexo 3 está a tabela contendo a localização de cada novo gene, bem como o seu tamanho, localização na fita sense ou antisense e em quais condições foi expresso.

Dos 340 genes hipotéticos considerados expressos, buscou-se saber a função. Porém muitos deles ainda possuem função desconhecida. Utilizando o KEGG (MAO *et al.*, 2005) para encontrar as funções gênicas, apenas 18 desses genes tiveram alguma função já catalogada pelo KEGG, retornando um número KO (KEGG Orthologia). Os genes novos e suas respectivas funções estão na tabela 09.

Tabela 09. Funções gênicas de 18 novos genes que tiveram determinada função retornada pelo KEGG.

Nº do gene	KO (KEGG Orthology)	Produto do Gene	Via(s) metabólica(s) que participa	Condições expressas
N_26	K03154	thiS, proteína transportadora de enxofre.	Sistema de revezamento de Enxofre	AltoO2, Controle, Nar, Nitrato, PS3
N_36	K09749	Proteína hipotética em mais de 350 organismos.		AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_73	K13028	oxdA, aldoxima desidratase		AD1, BaixoO2, Controle, Nitrato, PS3
N_81	K04062	osmB, lipoproteína osmoticamente induzida.		BaixoO2, Controle, Nar, Nitrato, PT
N_88	K01810	gpi, isomerase de glicose-6-fosfato	Metabolismo do carbono. Glicólise. Via das pentoses fosfato. Amido e metabolismo da sacarose.	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_95	K09791	Proteína hipotética em mais de mil organismos.		AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_101	K03640	pal, lipoproteína associada a peptidoglicano		AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_120	K03521	fixA, etfB, flavoproteína de transferência de elétrons subunidade beta.		AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_241	K13931	mdcC, malonato subunidade descarboxilase delta.		AltoO2, BaixoO2, Controle, Nitrato, PT, WR
N_245	K03424	TatD, Proteína família DNase		BaixoO2, Controle, Nitrato
N_250	K02014	TC.FEV.OM,		Nar, Nitrato

Nº do gene	KO (KEGG Orthology)	Produto do Gene	Via(s) metabólica(s) que participa	Condições expressas
		proteína receptora membrana externa complexo de ferro.		
N_262	K00242	sdhD, succinato desidrogenase subunidade âncora de membrana.	Metabolismo do carbono. Ciclo do Citrato. Metabolismo butanoato. Fosforilação oxidativa.	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_344	K02897	proteína grande subunidade ribossomal L25	Ribossomo.	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_353	K10439	rbsB, proteína de ligação de substrato, sistema de transporte ribose	Transportadores ABC. Quimiotaxia bacteriana.	Controle, Nar, Nitrato, PS1, PT, WR
N_378	K03745	SlyX proteína		AD3, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_382	K11746	KefF proteína, auxiliar sistema de potássio-efluxo regulado-glutationa		BaixoO2, Controle, Nar, Nitrato
N_410	K04096	Smf, Proteína de processamento de DNA.		AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_420	K03501	gidB, rsmG, 16S rRNA (guanina 527-N7)-metiltransferase		AD3, AltoO2, BaixoO2, Nar, Nitrato, PS1, PS3, PT

5.7.2 Identificação de ncRNA de *H. seropedicae*

É utilizado o termo RNA não-codificante (ncRNA) para o RNA que não codifica para uma proteína, mas isto não significa que tais RNAs não possuem função. Muitos ncRNAs foram confirmados com papéis reguladores em diversos processos biológicos. As tecnologias de sequenciamento em larga escala indicam a presença de transcritos que não codificam proteínas em diversos organismos procariotos e eucariotos (MATTICK & MAKUNIN, 2006 ; QU & ADELSON, 2012).

Tendo por base um conjunto de ncRNAs preditos por MORENO (2013) que utilizou programas que predizem promotores e terminadores, e também o programa Cufflinks que

identifica transcritos em regiões em que não há ORFs. Buscou-se observar a transcrição dos ncRNAs com expressão ainda não confirmada.

O critério para afirmar que um ncRNA estava expresso foi o mesmo seguido por MORENO (2013), que confirmou a expressão dos ncRNAs preditos usando os dados de RNA-Seq em três condições: Controle, Naringenina e Associado a planta. Confirmando um total de 98 ncRNAs.

Antes do alinhamento das leituras de RNA-seq, os nucleotídeos das regiões 5' - 3' foram aparados (*trimming*) para 35 nucleotídeos utilizando um *script* em linguagem Perl. Em seguida as sequências foram mapeadas no genoma da bactéria *H. seropedicae* SmR1 utilizando a ferramenta SHRiMP, com o seguinte comando de execução:

```
gmapper-cs arquivo.csfasta referencia.fasta -N8 -o1 -h80% -E 2>arquivo.log>arquivo.sam
```

Neste comando o parâmetro -N se refere ao número de núcleos de processamento que foi utilizado pelo computador, sendo então, 8 núcleos. O parâmetro -o determina o número de vezes que um *read* pode ser alinhado, neste caso, apenas 1 vez. O parâmetro -h estipula a porcentagem de homologia entre as sequências, sendo de 80%. E -E permite o arquivo de saída no formato sam.

O arquivo de mapeamento foi convertido para o modo binário (extensão .bam) com a seguinte sequência de comandos:

- 1) samtools view -bS arquivo.SAM > arquivo.BAM (Conversão do mapeamento do formato SAM para o formato BAM);
- 2) samtools sort arquivo.BAM arquivobam.sorted (Alinhamento das coordenadas do mapeamento) ;
- 3) samtools index arquivo.sorted.bam (Classifica o índice para que o acesso ao arquivo seja mais rápido).

Os ncRNAs candidatos que obtiveram o valor de cobertura igual ou superior a cinco, em pelo menos uma das condições de RNA-Seq, foram considerados como ncRNAs confirmados, representando um total de 71 novos ncRNAs. A tabela 10 mostra a quantidade de novos ncRNAs por condição. A tabela com a localização no genoma de todos os novos ncRNA confirmados estão no Anexo 4.

TABELA 10. Quantidade de ncRNAs confirmados em cada condição.

Condições	Cobertura ≥ 5
NFbHPN-Malato-Alto Oxigênio	28
NFbHPN-Malato-Baixo Oxigênio	31
NFbHP-Malato-Nitrato	48
Planctônico–milho-1 dia	29
Planctônico-milho-3dias	29
Aderido–milho- 1 dia	18
Aderido-milho3dias	30
Planctônico- trigo	41
Aderido-trigo	11

6. Conclusões

Os dados de transcriptoma (RNA-Seq) de *H. seropedicae* cultivado em 11 condições de crescimento diferentes (NFbHPN-Malato; NFbHPN-Malato-Naringenina; NFbHPN-Malato-Nitrato; Planctônico–milho-1dia; Planctônico-milho-3dias; Aderido–milho-1dia; Aderido-milho-3dias; Planctônico-trigo; Aderido-trigo; NFbHPN-Malato-Alto Oxigênio e NFbHPN-Malato-Baixo Oxigênio), permitiu a:

1 – Identificação de 583 genes que estão expressos em todas as 11 condições e que uma porcentagem desses genes pertence a vias metabólicas essenciais para a bactéria.

2 – Anotação de 340 novos genes em *H. seropedicae*.

3 – Identificação de 75 novos ncRNAs em *H. seropedicae*

4 – Identificação de 861 operons presentes em *H. seropedicae*.

A utilização de dados de RNAseq é uma ferramenta que pode ampliar o conhecimento sobre o panorama metabólico de *H. seropedicae* em diferentes condições e que também permite o aumento de informações geradas pelo genoma desta bactéria.

REFERÊNCIAS

- AIZAWA, S. I. **What is essencial for flagellar assembly?** In Ken F. Jarrell (Ed.). Pili and Flagella – Current Research and Future Trends. Norfolk, UK. 2009.
- ALDRIDGE, P.; HUGHES, K. T. Regulation of flagellar assembly. **Current Opinion in Microbiology**. Vol. 5, p. 160-165, 2002.
- ALTERTHUM, F. **Nutrição e metabolismo de bactérias**. In: Luiz Rachid Trabulsi (Ed.) e Flavio Alterthum (Ed.). Microbiologia. Editora Atheneu, 4ª edição, 2005, p. 20-30.
- BALDANI, J. I. et al. Characterization of *Herbaspirillum seropedicae*, a root-associated nitrogen fixing bacterium. **Intl. J. Sys. Bact.**, v. 36, p. 86-93, 1986.
- BALDANI, V. L. D., BALDANI, J. I., DÖBEREINER, J. Inoculation of rice plants with the endophytic diazotrophs *Herbaspirillum seropedicae* and *Bukholderia* spp. **Biol. Fertil. Soils**, 30, 485-491, 2000.
- BALSANELLI, E.; **Caracterização de fatores moleculares envolvidos na interação de Herbaspirillum seropedicae com gramíneas**. Tese de doutorado, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná. Curitiba, 2013.
- BALSANELLI, E.; TULESKI, T. R.; DE BAURA, V. A.; YATES, M. G.; CHUBATSU, L. S.; PEDROSA, F. O.; DE SOUZA, E. M.; MONTEIRO, R. A.; CASCALES, E. Maize Root Lectins Mediate the Interaction with *Herbaspirillum seropedicae* via N-Acetyl Glucosamine Residues of Lipopolysaccharides. **Plos One**, v. 8, p. e77001, 2013.
- BATISTA, M. B.; SFEIR, M. Z. T.; FAORO, H.; WASSEM, R.; STEFFENS, M. B. R.; PEDROSA, F. O.; SOUZA, E. M.; DIXON, R.; MONTEIRO, R. A. The *Herbaspirillum seropedicae* SmR1 Fnr orthologs controls the cytochrome composition of the electron transport chain. **Scientific Reports**, v. 3, p. 2544, 2013.
- BENNETZEN, J. L.; HALL, B. D. Codon selection in yeast. **Journal of Biological Chemistry** **257**: p. 3026-3031, 1982.

BODDEY, R. M.; DE OLIVEIRA, O.C.; URQUIAGA, S.; REIS, V.M.; OLOVARES, F.L.; BALDANI, V.L.D.; DOBEREINER, J. Biological nitrogen fixation associated with sugar cane and rice: Contribution and prospects for improvements. **Plant Soil**, 174: 195-209, 1995.

BONATO, P. **Genes do metabolismo de Nitrato em *Herbaspirillum seropedicae*: regulação transcricional e análise funcional**. Dissertação (Mestrado em Bioquímica) - Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, 2012.

BROUWER, R. W. W.; HIJUM, S. A. F. T.; KUIPERS, O. P.: MINOMICS: visualizing prokaryote transcriptomics and proteomics data in a genomic context. **Bioinformatics**, vol. 25, nº 1, 2008.

BROUWER, R. W. W.; KUIPERS, O. P.; HIJUM, S. A. F. T. V.: The relative value of operons predictions. **Briefings in Funcional Genomics**. V. 9 nº 5. P. 367-375. 2008.

BULLARD, J. H.; PURDOM, E.; HANSEN, K. D.; DUDOIT, S.: Evaluation of statistical methods for normalization and differential expression in mRNA -Seq experiments. **BMC Bioinformatics**, 2010.

CHU, Y; COREY, D. R. RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. **Nucleic Acid Therapeutics**, Volume 22, Number 4, 2012.

CLC BIO WORKBENCH (2012). Disponível em: <<http://clcbio.com/index.php?id=1240>> acessado em set/2012.

CODONW (2013) <<http://codonw.sourceforge.net>> (acessado em fev/2013)

CROUCHER, N. J.; THOMSON, N. R.: Studying bacterial transcriptomes using RNA-Seq. **Current Opinion in Microbiology**, 13(5): 619-624, 2010.

DIJK, J. P. V.; CANKAR, K.; HENDRIKSEN, P. J. M.; BEENEN, H. G.; ZHU, M.; SCHEFFER, S.; SHEPHERD, L. V. T.; STEWART, D.; DAVIES, H. V.; LEIFERT, C.; WILKOCKSON, S. J.; GRUDEN, K.; and KOK, E. J. The identification and Interpretation of Differences in the Transcriptomes of Organically and Conventionally Grow Potato Tubers. **Journal of Agricultural and Food Chemistry**. 60 (9), p. 2090–2101, 2012.

- DÖBEREINER, J.; PEDROSA, F. O. Nitrogen-fixing bacteria in nonleguminous crop plants. *Science Tech publishers/Springer-Verlag*, 1987.
- DU, Y.; ZHANG, L.; XU, F.; HUANG, B.; ZHANG, G.; LI, L. Validation of housekeeping genes as internal controls for studying gene expression during Pacific oyster (*Crassostrea gigas*) development by quantitative real-time PCR, *Fish & Shellfish Immunology*, Vol. 34, 2013, 939-945. <http://dx.doi.org/10.1016/j.fsi.2012.12.007> (25/07/2013).
- EISENBERG, E; LEVANON, E. Y. Human housekeeping genes are compact. *Trends in Genetics*, Volume 19, July 2003, Pages 362-365.
- FASTQC: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (acessado em dez. 2012)
- GREER, S.; HONEYWELL, R.; GELETU, M.; ARULANANDAM, R.; RAPTIS, L. Housekeeping genes; expression levels may change with density of cultured cells, *Journal of Immunological Methods*, Vol. 355, 2010, p. 76-79.
- GYANESHWAR, P.; JAMES, E. K.; REDDY, P. M.; LADHA, J. K. *Herbaspirillum* colonization increases growth and nitrogen accumulation in aluminium-tolerant rice varieties. *New Phytol.*, 154, 131-145, 2002.
- HOUSELEY, J.; TOLLERVEY, D. The many pathways of RNA degradation. *Cell* (review), vol. 136, p. 763-776, 2009.
- IKEMURA, T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* system. *Journal of Molecular Biology* **151**: p. 389-409, 1981.
- JAMES, E. K.; GYANESHWAR, P., MATHAN, N. BARRAQUIO, Q. L., REDDY, R. M., IANNETTA, P. P. M., *et. al.* Infection and colonization of rice seedlings by the plant growth-promoting bacterium *Herbaspirillum seropedicae* Z67. *Mol. Plant-Microbe Interact*, 15, 894-906, 2002.
- JAMES, E.K.; OLIVARES, F. L. Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs. *Crit. Rev. Plant Sci.*, 17: 77-119, 1998.
- KANEHISA, M.; GOTO, S.; SATO, Y.; KAWASHIMA, M.; FURUMICHI, M.; TANABE, M. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Research*, vol. 42, p. 199-205, 2013.

KEEG (2013) <<http://www.genome.jp/kegg>> (acessado em 11/03/2013).

LEHNINGER, A.L.; NELSON, D.L.; COX, M.M.; **Principles of Biochemistry**. 5ª Ed. 2008.

LI, H.; HANDSAKER, B.; WYSOKER, A.; FENNELL, T.; RUAN, J.; HOMER, N.; MARTH, G.; ABECASIS, G.; DURBIN, R. and 1000 Genome Project Data Processing Subgroup - The Sequence alignment/map (SAM) format and SAMtools. **Bioinformatics**, 25, 2078-9, (2009).

LI, S.; DONG, X. and SU, Z. Reconstruction of Operon Structures in Prokaryotes using Directional RNA-seq Short Reads by a Hidden Markov Model, 2014 (Submitted).

MAO, X. ; MA, Q.; ZHOU, C.; CHEN, X.; ZHANG, H.; YANG, J.; MAO, F.; LAI, W. and XU, Y. DOOR 2.0: presenting operons and their functions through dynamic and integrated views. **Nucleic Acids Research**, p. 1-6, 2013.

MAO, X.; CAI, T.; OLYARCHUK, J. G.; WEI, L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. **Bioinformatics**, vol21, nº19, 2005, p. 3787-3793.

MARQUES, M. V.; SILVA, A. M. Genômica Funcional: Transcriptoma. In: MIR, L. (Org). **Genômica**. São Paulo: Atheneu, 2004. P.119-135.

MATTICK, J. S.; MAKUNIN, I. V.: Non-coding RNA. **Human Molecular Genetics**, vol. 15, p. 17-29, 2006.

MCCLURE, R.; BALASUBRAMANIAN, D.; SUN, Y.; BOBROVSKYY, M.; SUMBY, P.; GENCO, C. A.; ANDERPOOL, C. K.; TJADEN, B.: computational analysis of bacterial RNA-Seq data. **Nucleic Acids Research**. p. 1-16, 2013.

MONTEIRO, R. A.; BALSANELLI, E.; WASSEM, R.; MARIN, A. M.; BRUSAMARELLO-SANTOS, L. C. C.; SCHMIDT, M. A.; TADRA-SFEIR, M. Z.; PANKIEVICZ, V. C. S.; CRUZ, L. M.; CHUBATSU, L. S.; PEDROSA, F. O.; SOUZA, E. M. Herbaspirillum-plant interactions: microscopical, histological and molecular aspects. **Plant and Soil**, v. xx, p. 01-xx, 2012.

MORENO, L. F., **Predição e análise de expressão de RNAs não codificadores com função regulatória presentes na bactéria *Herbaspirillum seropedicae* Smr1.** Dissertação (Mestrado em Bioquímica) – Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, 2013.

MORIYA, Y., ITOH, M., OKUDA, S., YOSHIZAWA, A., and KANEHISA, M.; KAAS: an automatic genome annotation and pathway reconstruction server. **Nucleic Acids Res.** 35, p.182-185, 2007.

MORTAZAVI, A., B. A. WILLIAMS, *et al.* (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. **Nat Methods** 5(7): 621-628.

OLIVER, H. F.; ORSI, R. H.; PONNALA, L.; KEICH, U.; WANG, W.; SUN, Q.; CARTINHO, S. W.; FILIATRAULT, M. J.; WEIDMANN, M.; e BOOR, K. J. Deep RNA sequencing of *L. monocytogenes* reveals overlapping and extensive stationary phase and sigma B-dependent transcriptomes, including multiple highly transcribed noncoding RNAs. **BMC Genomics**, 10, 2009.

OSHLACK, A.; WAKEFIELD, M. J. Transcript length bias in RNA-seq data confounds systems biology. **BioMed Central**, 2009.

PANKIEVICZ, V. C. S. **Investigação de mecanismos envolvidos na interação entre *Herbaspirillum seropedicae* e gramíneas.** Projeto de Tese de doutorado, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná. Curitiba, 2013. (não publicada).

PEDROSA, F.O.; MONTEIRO, R.A.; WASSEM, R.; CRUZ, L.M.; AYUB, R.A., *et al.* (2011) Genome of *Herbaspirillum seropedicae* Strain SmR1, a Specialized Diazotrophic Endophyte of Tropical Grasses. **PLoS Genetics** 7(5): e1002064. doi:10.1371/journal.pgen.1002064

PERKINS, T. T.; KINGSLEY, R. A.; FOOKES, M. C.; GARDNER, P. P.; JAMES, K. D.; YU, L.; ASSEFA, S. A.; HE, M.; CROUCHER, N. J.; PICKARD, D. J.; MASKELL, D. J.; PARKHILL, J.; CHOUDHARY, J.; THOMSON, N. R.; DOUGAN, G. A Strand-Specific RNA-Seq Analysis of the Typhoid Bacillus *Salmonella* Typhi. **Plos Genetics**, 2009.

QU, Z.; ADELSON, D. L.: Evolucinary conservation and funcional roles of ncRNA. **Frontiers in Genetics**, Vol. 3, 2012.

RAITTZ, R. T. A new hybrid bioinformatics platform for the identification of proteins coding regions in prokaryotes. **Abstracts Book: 1st International Workshop on Bioinformatics**. 2010. Disponível em: www.bioinfo.ufpr.br/iwb

RAMOS, R.T.J.; CARNEIRO, A. R.; BAUMBACH, J; AZEVEDO, V.; SCHNEIDER, P. C.; SILVA, A. Analysis of quality raw data of second generation sequences with Quality Assessment Software. **BMC Research Notes**, 2011 4:130.

RUMBLE, S. M.; LACROUTE, P.; DALCA, A.V., FIUME, M., SIDOW, A., et al. SHRiMP: Accurate Mapping of Short Color-space Reads. **PLoS Comput Biol** 5(5), 2009.

RUTHERFORD, K; PARKHILL, J; CROOK, J; HORSNELL, T; RICE, P; RAJANDREAM, M. A.; BARRELL, B. Artemis: sequence visualization and annotation. **Bioinformatics** (Oxford, England) v. 16; nº 10; p. 944-5. 2000.

SAMTOOLS (2012). <<http://samtools.sourceforge.net>> acessado em set/2012.

SANGER (2012). Disponível em: <<http://www.sanger.ac.uk/resources/software/artemis>> acessado em set/2012.

SHARP, P. M., and LI, W. H. The codon adaptation index a measure of directional synonymous codon usage bias, and its potential applications. **Nucleic Acids Research** 15: p. 1281-1295, 1987.

SOLOVYEV, V.; SALAMOV A. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies* (Ed. R.W. Li), **Nova Science Publishers**, 2011, p. 61-78

SOREK, R; COSSART, P. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. **Nature Reviews Genetics**, v. 11, p. 9-16, 2010.

TABOADA, B.; CIRIA, R.; MARTINEZ-GUERRER, C. E.; MERINO, E. ProOpDB: Prokaryotic Operon DataBase, **Nucleic Acids Research**, 40(D1), 2012.

TADRA-SFEIR, M.Z. **Caracterização funcional de genes de *Herbaspirillum seropedicae* regulados pelo flavonóide naringenina**. Tese de doutorado, Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná. Curitiba, 2011.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nature Reviews Genetics**, v. 10, p. 57-63, 2009.

ANEXO 1 - Lista dos genes expressos nas 11 condições de transcriptoma com RPKM ≥ 100 OU Cobertura ≥ 2 .

Hsero_0001, Hsero_0002, Hsero_0003, Hsero_0004, Hsero_0005, Hsero_0063, Hsero_0068, Hsero_0082, Hsero_0083, Hsero_0085, Hsero_0086, Hsero_0087, Hsero_0088, Hsero_0089, Hsero_0097, Hsero_0099, Hsero_0100, Hsero_0101, Hsero_0102, Hsero_0103, Hsero_0104, Hsero_0105, Hsero_0106, Hsero_0109, Hsero_0110, Hsero_0111, Hsero_0112, Hsero_0113, Hsero_0114, Hsero_0115, Hsero_0116, Hsero_0117, Hsero_0118, Hsero_0119, Hsero_0120, Hsero_0121, Hsero_0122, Hsero_0123, Hsero_0124, Hsero_0125, Hsero_0126, Hsero_0127, Hsero_0128, Hsero_0129, Hsero_0130, Hsero_0131, Hsero_0132, Hsero_0133, Hsero_0134, Hsero_0135, Hsero_0136, Hsero_0137, Hsero_0138, Hsero_0139, Hsero_0140, Hsero_0141, Hsero_0144, Hsero_0165, Hsero_0167, Hsero_0168, Hsero_0174, Hsero_0185, Hsero_0197, Hsero_0199, Hsero_0239, Hsero_0303, Hsero_0305, Hsero_0318, Hsero_0326, Hsero_0327, Hsero_0328, Hsero_0331, Hsero_0335, Hsero_0340, Hsero_0341, Hsero_0343, Hsero_0346, Hsero_0353, Hsero_0363, Hsero_0364, Hsero_0365, Hsero_0366, Hsero_0369, Hsero_0372, Hsero_0374, Hsero_0377, Hsero_0380, Hsero_0381, Hsero_0382, Hsero_0388, Hsero_0389, Hsero_0390, Hsero_0415, Hsero_0423, Hsero_0426, Hsero_0431, Hsero_0453, Hsero_0464, Hsero_0467, Hsero_0468, Hsero_0482, Hsero_0483, Hsero_0486, Hsero_0497, Hsero_0501, Hsero_0505, Hsero_0608, Hsero_0614, Hsero_0615, Hsero_0616, Hsero_0633, Hsero_0634, Hsero_0646, Hsero_0647, Hsero_0648, Hsero_0655, Hsero_0657, Hsero_0658, Hsero_0670, Hsero_0671, Hsero_0672, Hsero_0677, Hsero_0680, Hsero_0681, Hsero_0682, Hsero_0687, Hsero_0688, Hsero_0698, Hsero_0705, Hsero_0741, Hsero_0760, Hsero_0761, Hsero_0837, Hsero_0844, Hsero_0858, Hsero_0859, Hsero_0860, Hsero_0866, Hsero_0867, Hsero_0869, Hsero_0870, Hsero_0875, Hsero_0876, Hsero_0882, Hsero_0883, Hsero_0921, Hsero_0964, Hsero_0996, Hsero_0997, Hsero_1053, Hsero_1104, Hsero_1121, Hsero_1122, Hsero_1165, Hsero_1171, Hsero_1181, Hsero_1182, Hsero_1183, Hsero_1200, Hsero_1214, Hsero_1216, Hsero_1223, Hsero_1236, Hsero_1237, Hsero_1238, Hsero_1376, Hsero_1377, Hsero_1379, Hsero_1382, Hsero_1383, Hsero_1384, Hsero_1393, Hsero_1396, Hsero_1397, Hsero_1398, Hsero_1399, Hsero_1404, Hsero_1417, Hsero_1423, Hsero_1425, Hsero_1427, Hsero_1429, Hsero_1434, Hsero_1440, Hsero_1445, Hsero_1447, Hsero_1453, Hsero_1454, Hsero_1455, Hsero_1465, Hsero_1467, Hsero_1468, Hsero_1503, Hsero_1506, Hsero_1507, Hsero_1508, Hsero_1509, Hsero_1547, Hsero_1548, Hsero_1549, Hsero_1623, Hsero_1624, Hsero_1626, Hsero_1632, Hsero_1633, Hsero_1639, Hsero_1662, Hsero_1684, Hsero_1716, Hsero_1719, Hsero_1727, Hsero_1729, Hsero_1734, Hsero_1735, Hsero_1736, Hsero_1737, Hsero_1754, Hsero_1755, Hsero_1765, Hsero_1767, Hsero_1768, Hsero_1769, Hsero_1770, Hsero_1771, Hsero_1772, Hsero_1773, Hsero_1774, Hsero_1775, Hsero_1776, Hsero_1777, Hsero_1778, Hsero_1779, Hsero_1780, Hsero_1785, Hsero_1802, Hsero_1866, Hsero_1869, Hsero_1870, Hsero_1872, Hsero_1877, Hsero_1885, Hsero_1889, Hsero_1895, Hsero_1920, Hsero_1921, Hsero_1924, Hsero_1926, Hsero_1927, Hsero_1928, Hsero_1929, Hsero_1930, Hsero_1931, Hsero_1934, Hsero_1935, Hsero_1936, Hsero_1957, Hsero_1976, Hsero_1977, Hsero_1978, Hsero_1979, Hsero_1980, Hsero_1983, Hsero_1984, Hsero_2020, Hsero_2023, Hsero_2024, Hsero_2026, Hsero_2027, Hsero_2029, Hsero_2030, Hsero_2031, Hsero_2033, Hsero_2035,

Hsero_2036, Hsero_2037, Hsero_2038, Hsero_2039, Hsero_2040, Hsero_2041, Hsero_2042,
 Hsero_2043, Hsero_2051, Hsero_2052, Hsero_2053, Hsero_2054, Hsero_2055, Hsero_2057,
 Hsero_2058, Hsero_2059, Hsero_2060, Hsero_2061, Hsero_2062, Hsero_2065, Hsero_2066,
 Hsero_2067, Hsero_2069, Hsero_2070, Hsero_2071, Hsero_2074, Hsero_2075, Hsero_2076,
 Hsero_2077, Hsero_2078, Hsero_2080, Hsero_2082, Hsero_2083, Hsero_2085, Hsero_2087,
 Hsero_2110, Hsero_2121, Hsero_2128, Hsero_2129, Hsero_2141, Hsero_2149, Hsero_2154,
 Hsero_2162, Hsero_2165, Hsero_2178, Hsero_2179, Hsero_2180, Hsero_2181, Hsero_2182,
 Hsero_2186, Hsero_2187, Hsero_2195, Hsero_2203, Hsero_2205, Hsero_2206, Hsero_2215,
 Hsero_2242, Hsero_2243, Hsero_2244, Hsero_2290, Hsero_2293, Hsero_2353, Hsero_2354,
 Hsero_2356, Hsero_2462, Hsero_2463, Hsero_2515, Hsero_2538, Hsero_2569, Hsero_2580,
 Hsero_2581, Hsero_2582, Hsero_2583, Hsero_2618, Hsero_2622, Hsero_2725, Hsero_2770,
 Hsero_2771, Hsero_2772, Hsero_2785, Hsero_2876, Hsero_2904, Hsero_2905, Hsero_2912,
 Hsero_2920, Hsero_2926, Hsero_2927, Hsero_2928, Hsero_2929, Hsero_2930, Hsero_2942,
 Hsero_2943, Hsero_2945, Hsero_2946, Hsero_2947, Hsero_2948, Hsero_2949, Hsero_2952,
 Hsero_2953, Hsero_2955, Hsero_2956, Hsero_2957, Hsero_2959, Hsero_2967, Hsero_2968,
 Hsero_2969, Hsero_2970, Hsero_2971, Hsero_2972, Hsero_2973, Hsero_2974, Hsero_2975,
 Hsero_2976, Hsero_2979, Hsero_2981, Hsero_2982, Hsero_2983, Hsero_2984, Hsero_2985,
 Hsero_2986, Hsero_2988, Hsero_2989, Hsero_2997, Hsero_3004, Hsero_3022, Hsero_3028,
 Hsero_3029, Hsero_3063, Hsero_3081, Hsero_3083, Hsero_3096, Hsero_3127, Hsero_3134,
 Hsero_3135, Hsero_3142, Hsero_3143, Hsero_3144, Hsero_3145, Hsero_3157, Hsero_3196,
 Hsero_3197, Hsero_3198, Hsero_3199, Hsero_3201, Hsero_3202, Hsero_3215, Hsero_3228,
 Hsero_3266, Hsero_3272, Hsero_3275, Hsero_3276, Hsero_3280, Hsero_3284, Hsero_3288,
 Hsero_3290, Hsero_3293, Hsero_3296, Hsero_3297, Hsero_3298, Hsero_3299, Hsero_3360,
 Hsero_3413, Hsero_3419, Hsero_3424, Hsero_3425, Hsero_3449, Hsero_3451, Hsero_3455,
 Hsero_3465, Hsero_3471, Hsero_3472, Hsero_3473, Hsero_3495, Hsero_3499, Hsero_3500,
 Hsero_3511, Hsero_3556, Hsero_3572, Hsero_3577, Hsero_3582, Hsero_3586, Hsero_3601,
 Hsero_3604, Hsero_3605, Hsero_3606, Hsero_3608, Hsero_3612, Hsero_3614, Hsero_3680,
 Hsero_3688, Hsero_3689, Hsero_3690, Hsero_3695, Hsero_3696, Hsero_3727, Hsero_3731,
 Hsero_3749, Hsero_3780, Hsero_3782, Hsero_3789, Hsero_3833, Hsero_3834, Hsero_3845,
 Hsero_3846, Hsero_3847, Hsero_3848, Hsero_3849, Hsero_3850, Hsero_3854, Hsero_3857,
 Hsero_3858, Hsero_3880, Hsero_3906, Hsero_3911, Hsero_3918, Hsero_3920, Hsero_3923,
 Hsero_3924, Hsero_3929, Hsero_3932, Hsero_3957, Hsero_3961, Hsero_3968, Hsero_3973,
 Hsero_3974, Hsero_3975, Hsero_4035, Hsero_4045, Hsero_4046, Hsero_4047, Hsero_4059,
 Hsero_4066, Hsero_4075, Hsero_4076, Hsero_4079, Hsero_4093, Hsero_4141, Hsero_4150,
 Hsero_4151, Hsero_4168, Hsero_4169, Hsero_4172, Hsero_4188, Hsero_4193, Hsero_4194,
 Hsero_4195, Hsero_4196, Hsero_4197, Hsero_4198, Hsero_4199, Hsero_4200, Hsero_4201,
 Hsero_4202, Hsero_4203, Hsero_4204, Hsero_4205, Hsero_4206, Hsero_4207, Hsero_4208,
 Hsero_4209, Hsero_4210, Hsero_4211, Hsero_4212, Hsero_4213, Hsero_4214, Hsero_4215,
 Hsero_4216, Hsero_4217, Hsero_4218, Hsero_4219, Hsero_4220, Hsero_4221, Hsero_4222,
 Hsero_4227, Hsero_4252, Hsero_4253, Hsero_4271, Hsero_4272, Hsero_4289, Hsero_4295,
 Hsero_4296, Hsero_4301, Hsero_4309, Hsero_4310, Hsero_4325, Hsero_4336, Hsero_4339,
 Hsero_4340, Hsero_4349, Hsero_4362, Hsero_4363, Hsero_4364, Hsero_4365, Hsero_4366,

Hsero_4367, Hsero_4368, Hsero_4369, Hsero_4412, Hsero_4420, Hsero_4430, Hsero_4662, Hsero_4671, Hsero_4677, Hsero_4678, Hsero_4707, Hsero_4771, Hsero_4775, Hsero_4789, Hsero_4790, Hsero_4796,

Genes Adicionados:

Glicolise: Hsero_1099, Hsero_0163, Hsero_1097, Hsero_0472, Hsero_3549;

Montagem do flagelo: Hsero_2034, Hsero_2025, Hsero_2056, Hsero_2049, Hsero_2032;

Ciclo de Krebs: Hsero_1543;

ANEXO 2 - 861 operons retornados pelo *software* Rockhopper.

Início	Fim	Strand	Nº de genes	Genes
23342	25449	+	2	czcR, czcS
29277	34816	+	3	Hsero_0020, Hsero_0021, Hsero_0022
34891	37553	+	2	Hsero_0023, Hsero_0024
37678	39643	+	3	Hsero_0025, Hsero_0026, Hsero_0027
42910	44049	-	2	Hsero_0031, Hsero_0032
46587	48159	-	2	fecR, fecI
63766	66169	+	3	Hsero_0049, Hsero_0050, feoB
69098	71909	-	3	tauC, tauB, tauA
74416	76379	+	2	Hsero_0061, speB
77496	78976	+	2	glnP, glnQ
79037	80577	-	2	Hsero_0066, Hsero_0067
82258	86458	-	3	ntrY, Hsero_0071, rrmB
89932	91511	-	2	cdsA, plsC
92009	95807	-	3	Hsero_0078, Hsero_0079, Hsero_0080
98251	100938	+	3	Hsero_0083, glnK, amtB
101286	104079	+	3	gshA, gshB, pts
106756	108494	+	2	metX, Hsero_0092
111797	112769	+	2	secE, nusG
112930	114056	+	2	rplK, rplA
124609	128080	+	2	cirA, hemS
133522	138307	+	10	rplC, rplD, rplW, rplB, rpsS, rplV, rpsC, rplP, rpmC, rpsQ
138504	144706	+	14	rplN, rplX, rplE, rpsN, rpsH, rplF, rplR, rpsE, rpmD, rplO, secY, infA, rpmJ, rpsM
147754	149942	+	2	cutA, dsbD
156854	158874	-	2	rpoE, hemB
166667	168133	-	2	gpsA, Hsero_0159
168798	169468	-	2	grxC, Hsero_0162
170458	173370	+	2	yibP, prc
181936	183900	+	2	Hsero_0175, Hsero_0176
185026	187023	+	2	livM, livG
192695	194267	+	2	Hsero_0183, Hsero_0184
200180	201089	-	2	gspG, cyaY
206791	209972	+	2	Hsero_0195, otsA
210152	211162	+	2	Hsero_0197, Hsero_0198
213343	215382	-	2	Hsero_0200, Hsero_0201
215425	219103	-	5	Hsero_0202, Hsero_0203, Hsero_0204, Hsero_0205, Hsero_0206
220331	223091	-	4	Hsero_0209, Hsero_0210, Hsero_0211, Hsero_0212
223160	227973	-	5	Hsero_0213, Hsero_0214, Hsero_0215, Hsero_0216,

Inicio	Fim	Strand	Nº de genes	Genes
				Hsero_0217
228457	230679	-	3	Hsero_0219, Hsero_0220, Hsero_0221
232038	236407	+	5	Hsero_0224, Hsero_0225, Hsero_0226, exeA, Hsero_0228
236489	238032	+	4	Hsero_0229, Hsero_0230, Hsero_0231, Hsero_0232
238395	240934	+	2	Hsero_0233, Hsero_0234
242567	244530	+	2	Hsero_0237, Hsero_0238
246845	250174	-	4	Hsero_0241, Hsero_0242, Hsero_0243, lldD
252738	254993	-	2	Hsero_0247, Hsero_0248
255099	259957	-	2	Hsero_0249, Hsero_0250
261271	263781	+	2	Hsero_0253, Hsero_0254
276107	279383	+	3	pta, ackA, phaC
280207	285958	+	3	Hsero_0267, Hsero_0268, Hsero_0269
286019	287832	-	2	Hsero_0270, Hsero_0271
296150	302528	+	3	Hsero_0279, Hsero_0280, Hsero_0281
302588	304238	+	2	Hsero_0282, Hsero_0283
308723	310630	-	2	Hsero_0289, Hsero_0290
313202	314319	-	2	Hsero_0292, exbD
325445	327225	+	2	tktA1, tktA2
328779	335539	+	3	Hsero_0306, Hsero_0307, Hsero_0308
342527	347290	-	2	Hsero_0314, Hsero_0315
347390	349598	-	2	Hsero_0316, Hsero_0317
359967	367535	+	7	mraZ, mraW, ftsL, ftsI, murE, murF, mraY
367630	376572	+	7	murD, ftsW, murG, murC, ddlB, ftsQ, ftsA
384760	388951	+	5	Hsero_0347, Hsero_0348, argJ, Hsero_0350, mutT
388988	389958	-	2	Hsero_0352, Hsero_0353
390111	391606	-	2	coaE, pilD
391658	394025	-	2	pilC, pilB
394189	395748	-	2	corB, Hsero_0359
401333	402501	-	2	Hsero_0368, rppH
407919	410530	+	3	Hsero_0374, Hsero_0375, Hsero_0376
410664	412384	+	3	Hsero_0377, Hsero_0378, ampD
421059	423266	-	2	Hsero_0386, prmA
425925	427087	-	2	tlpA, Hsero_0392
428446	430480	+	2	cheD1, Hsero_0395
430556	436150	+	5	mpl, Hsero_0397, ubiC, vacB, tonB
436239	437788	+	2	aroE, mtgA
442429	444276	+	2	tyrS, dtd
444363	445664	+	2	Hsero_0409, gpmB
445806	447436	+	2	hipB, hipA
452144	453866	+	2	phrB, Hsero_0418
455531	457204	+	2	ruvA, ruvB

Inicio	Fim	Strand	Nº de genes	Genes
457384	458701	+	2	Hsero_0423, Hsero_0424
464921	467233	+	3	thiG, thiD, thiE
479558	482078	+	3	Hsero_0441, Hsero_0442, Hsero_0443
484958	486911	+	2	xylB, Hsero_0447
488946	490717	-	2	Hsero_0449, thrB
502854	506991	+	4	hcaE, tauB, tauC, ssnA
507263	509706	+	4	Hsero_0464, Hsero_0465, pyrR, pyrB
509882	511452	+	2	Hsero_0468, pyrX
515907	518062	+	2	Hsero_0474, kdtA
525480	531300	-	3	Hsero_0482, acrB, acrA
533252	534260	-	2	Hsero_0487, pcm
541924	545370	-	4	Hsero_0498, Hsero_0499, Hsero_0500, Hsero_0501
554479	556505	-	2	Hsero_0508, Hsero_0509
566985	568061	-	2	Hsero_0519, Hsero_0520
568141	569331	-	2	Hsero_0521, Hsero_0522
587597	589013	-	2	gloA, Hsero_0536
592577	595680	-	2	Hsero_0539, Hsero_0540
597730	600487	-	2	Hsero_0543, Hsero_0544
600817	602435	+	2	Hsero_0545, Hsero_0546
615746	617784	+	2	Hsero_0559, metE
619344	621131	+	2	Hsero_0563, Hsero_0564
625155	629906	-	3	Hsero_0569, ndeD, Hsero_0571
634325	635916	-	2	Hsero_0577, Hsero_0578
643002	644275	-	2	fabG, Hsero_0585
650901	655159	-	3	holA, rlpB, leuS
656765	661383	+	5	Hsero_0598, Hsero_0599, potC, potA, Hsero_0602
661659	663164	-	3	Hsero_0603, exbD, exbB
663324	664689	-	2	dapB, smpA
675786	677262	-	2	panB, Hsero_0618
677904	679187	+	2	Hsero_0619, cheY
679305	683606	+	3	Hsero_0621, Hsero_0622, Hsero_0623
683657	687471	+	4	cheW, cheR, cheB, Hsero_0627
687520	691013	-	4	folK, pcnB, serB, hda
694571	697001	-	2	Hsero_0635, Hsero_0636
703613	707060	+	2	Hsero_0641, Hsero_0642
707081	710580	-	3	Hsero_0643, Hsero_0644, Hsero_0645
714050	715895	-	2	Hsero_0649, Hsero_0650
717026	719556	-	3	Hsero_0652, Hsero_0653, Hsero_0654
728422	730015	-	2	Hsero_0663, Hsero_0664
731951	733056	+	2	azlC, Hsero_0669
734597	737121	+	2	pykA, fbaA
737213	740872	+	4	purC, purE, purK, sua5

Inicio	Fim	Strand	Nº de genes	Genes
752891	754038	-	2	exbD, exbB
754092	755381	-	3	tonB, bfd, Hsero_0693
755883	757526	-	2	trmU, Hsero_0695
759325	761075	+	2	pntAb, pntB
763194	766295	-	3	pepP, Hsero_0703, Hsero_0704
766393	770207	+	2	ostA, surA
770303	772134	+	2	pdxA, ksgA
782859	785386	-	2	Int, corC
787626	790564	-	3	Hsero_0721, phoH, miaB
794544	798348	+	3	Hsero_0728, Hsero_0729, Hsero_0730
798392	800377	-	2	Hsero_0731, Hsero_0732
802303	802962	-	2	Hsero_0734, Hsero_0735
803061	808097	-	3	Hsero_0736, Hsero_0737, Hsero_0738
809942	812465	+	3	Hsero_0741, Hsero_0742, Hsero_0743
812608	818708	+	4	Hsero_0744, Hsero_0745, Hsero_0746, clpB
821697	825866	+	2	Hsero_0749, Hsero_0750
827071	829105	+	2	Hsero_0752, Hsero_0753
831617	839148	-	5	Hsero_0756, Hsero_0757, Hsero_0758, Hsero_0759, Hsero_0760
841882	846759	+	4	Hsero_0763, Hsero_0764, Hsero_0765, Hsero_0766
846883	849568	+	2	Hsero_0767, Hsero_0768
863448	864894	+	4	Hsero_0782, Hsero_0783, Hsero_0784, Hsero_0785
867213	868645	+	2	Hsero_0787, Hsero_0788
868739	874196	+	6	hrpQ, hrcN, Hsero_0791, Hsero_0792, hrcQ, hrcR
874262	876981	+	4	hrcS, hrpX, hrcU, Hsero_0798
877060	879002	+	3	hrpB, hrcJ, hrpD
879871	882600	+	3	Hsero_0803, Hsero_0804, hrcC
883186	890145	+	6	Hsero_0806, pilN, pilO, pilP, gspE, bfpE
890862	893237	+	3	Hsero_0813, pilV, Hsero_0815
895179	897193	+	2	Hsero_0818, Hsero_0819
910154	915144	-	2	Hsero_0830, Hsero_0831
915600	918431	-	2	miaA, mutL
925067	926877	-	2	amiC, Hsero_0839
928103	929599	-	2	Hsero_0841, dctR
933287	935124	+	2	dctQ, dctM
936348	937817	+	2	ogT, xerD
937896	939045	+	2	Hsero_0850, Hsero_0851
951188	957435	+	2	Hsero_0861, uvrD
962832	967692	+	3	dnaG, Hsero_0869, rpoD
974612	975295	-	3	Hsero_0879, Hsero_0880, Hsero_0881
976834	981228	-	3	Hsero_0884, Hsero_0885, Hsero_0886
982298	989295	-	9	Hsero_0889, Hsero_0890, Hsero_0891, Hsero_0892,

Início	Fim	Strand	Nº de genes	Genes
				Hsero_0893, Hsero_0894, Hsero_0895, Hsero_0896, Hsero_0897
989443	992890	-	8	Hsero_0898, Hsero_0899, Hsero_0900, Hsero_0901, Hsero_0902, Hsero_0903, Hsero_0904, Hsero_0905
992978	994740	-	2	Hsero_0906, Hsero_0907
995776	998824	+	3	Hsero_0909, Hsero_0910, Hsero_0911
999952	1002531	-	5	Hsero_0914, Hsero_0915, Hsero_0916, Hsero_0917, Hsero_0918
1002956	1004715	+	2	Hsero_0920, Hsero_0921
1004985	1006417	-	2	Hsero_0922, Hsero_0923
1009108	1011495	+	2	atoB, Hsero_0927
1012983	1015955	-	3	oxdA, Hsero_0930, nthB
1028934	1034361	+	4	Hsero_0941, Hsero_0942, Hsero_0943, Hsero_0944
1044065	1046797	+	3	livG, livH, livM
1048249	1050271	+	2	Hsero_0956, livF
1051448	1053844	+	2	mmsB, Hsero_0960
1061225	1065407	+	5	Hsero_0967, Hsero_0968, Hsero_0969, Hsero_0970, Hsero_0971
1084227	1087575	-	2	Hsero_0985, fbpC
1094540	1096220	-	2	Hsero_0993, Hsero_0994
1103337	1109749	+	2	fhaB, fhaC
1111065	1114128	+	3	Hsero_1003, Hsero_1004, Hsero_1005
1114173	1115674	+	2	Hsero_1006, Hsero_1007
1115752	1120278	+	5	Hsero_1008, Hsero_1009, Hsero_1010, Hsero_1011, Hsero_1012
1123508	1126116	-	2	Hsero_1016, Hsero_1017
1127262	1129717	-	3	Hsero_1019, rbsK, rbsD
1129918	1133357	-	2	Hsero_1022, deoC
1139550	1142168	-	3	dgoAa, dgoK, galM
1143757	1147402	+	3	Hsero_1034, Hsero_1035, Hsero_1036
1160755	1161901	+	2	Hsero_1047, Hsero_1048
1162060	1164703	+	2	Hsero_1049, Hsero_1050
1167539	1172721	+	6	Hsero_1054, Hsero_1055, cfa, Hsero_1057, Hsero_1058, Hsero_1059
1175728	1178224	+	2	Hsero_1063, rbsC
1179050	1181436	+	2	araB, Hsero_1067
1182029	1183845	+	2	metX, Hsero_1070
1189835	1192149	-	2	gabD1, potC
1240197	1242057	+	2	idnO, idnD
1246224	1248588	-	2	Hsero_1097, zwf
1250683	1253180	+	2	edd, eda
1255837	1257683	-	2	tktC, tktB
1257724	1260623	-	2	dctQM, dctP
1262870	1265257	-	2	Hsero_1111, Hsero_1112

Inicio	Fim	Strand	Nº de genes	Genes
1268220	1269997	-	2	Hsero_1114, Hsero_1115
1270200	1274903	+	4	pcaK, hmgA, mhpD, gst
1284135	1285969	-	2	serA, Hsero_1128
1289247	1293023	+	3	Hsero_1131, Hsero_1132, Hsero_1133
1293254	1295315	+	2	Hsero_1134, dadA
1296866	1299575	+	2	betB, hipO
1309872	1312207	-	2	Hsero_1147, Hsero_1148
1327900	1330599	+	3	Hsero_1162, Hsero_1163, dadA
1333170	1334219	+	2	exbB2, exbD2
1342022	1342938	+	2	Hsero_1175, Hsero_1176
1343725	1345565	+	2	sun, Hsero_1179
1348557	1348985	-	2	rpmG, rpmB
1353110	1355848	+	3	livM, livG, livF
1364872	1365799	-	2	Hsero_1205, Hsero_1206
1367388	1369138	+	2	nasE, nasD
1376086	1379488	-	4	Hsero_1217, tolA, exbD, tolQ
1388011	1390187	+	2	tsaA, tgt
1390210	1392643	-	2	vsrB, corR
1392802	1396970	-	3	Hsero_1232, Hsero_1233, Hsero_1234
1413166	1414820	+	2	Hsero_1247, Hsero_1248
1427253	1429926	+	3	pstC, pstA, pstB
1430101	1431577	+	2	phoU, phoB
1442586	1444078	+	2	Hsero_1271, nahD
1444068	1445404	-	3	Hsero_1273, Hsero_1274, Hsero_1275
1458931	1460014	+	2	Hsero_1286, Hsero_1287
1461855	1462293	+	2	Hsero_1289, Hsero_1290
1463867	1474428	+	3	hlyB, Hsero_1294, Hsero_1295
1474494	1478437	+	2	Hsero_1296, Hsero_1297
1479980	1483545	+	2	Hsero_1300, Hsero_1301
1485791	1486874	-	2	catD, catC
1490397	1494592	+	5	cbeA, cbeB, cbeC, cbeD, yjgH
1496382	1500394	+	3	benE, Hsero_1315, Hsero_1316
1501173	1503204	+	3	Hsero_1318, Hsero_1319, mopI
1503226	1506629	-	2	Hsero_1321, iorA
1507031	1509725	+	3	Hsero_1323, Hsero_1324, Hsero_1325
1516212	1518320	+	2	Hsero_1331, xdhC
1527073	1528865	-	2	hcaA2, hcaA1
1535799	1536916	+	2	Hsero_1347, fabG
1538307	1543856	+	5	Hsero_1350, Hsero_1351, Hsero_1352, Hsero_1353, Hsero_1354
1545436	1551916	+	4	Hsero_1357, Hsero_1358, tolC, Hsero_1360
1555348	1558933	+	2	Hsero_1363, Hsero_1364

Inicio	Fim	Strand	Nº de genes	Genes
1559281	1562847	-	2	Hsero_1365, Hsero_1366
1569333	1570901	-	2	Hsero_1373, Hsero_1374
1573710	1578099	+	2	carA, carB
1585859	1587909	-	2	cccA, Hsero_1388
1600152	1602766	+	2	clpS, clpA
1602936	1604494	-	2	Hsero_1400, dut
1604568	1606311	-	2	coaBC, lspA
1610570	1615577	-	6	Hsero_1406, Hsero_1407, Hsero_1408, Hsero_1409, Hsero_1410, Hsero_1411
1621064	1623600	+	2	tktA, Hsero_1416
1634038	1636330	+	2	Hsero_1422, tldD
1639760	1641446	-	2	Hsero_1427, argG
1644953	1646244	-	2	Hsero_1431, Hsero_1432
1649226	1650827	+	2	thrB, tag
1666013	1666903	-	2	Hsero_1448, holC
1675515	1679000	+	2	aceF, lpdA
1681730	1683855	-	2	Hsero_1458, Hsero_1459
1685176	1686594	-	3	Hsero_1461, Hsero_1462, Hsero_1463
1690278	1691654	-	2	Hsero_1468, ttuC
1698724	1700762	+	2	fdsG, fdsB
1710520	1712280	-	2	tauB, tauA
1714024	1715129	-	3	Hsero_1490, Hsero_1491, Hsero_1492
1720081	1723840	+	5	Hsero_1496, Hsero_1497, Hsero_1498, Hsero_1499, Hsero_1500
1725186	1726873	+	2	Hsero_1502, etfA
1738777	1740917	-	2	envZ, ompR
1742253	1751608	+	4	Hsero_1520, Hsero_1521, Hsero_1522, Hsero_1523
1753733	1757016	+	3	Hsero_1525, Hsero_1526, corA
1757275	1764299	+	5	cirA, Hsero_1529, Hsero_1530, Hsero_1531, Hsero_1532
1764311	1766335	-	2	Hsero_1533, Hsero_1534
1776071	1779223	-	3	murI, fumA, Hsero_1544
1782149	1783422	+	2	Hsero_1547, Hsero_1548
1787738	1788385	-	2	Hsero_1554, cheY
1793752	1795448	-	2	Hsero_1559, Hsero_1560
1798181	1803590	+	5	Hsero_1563, dppB2, dppC2, dppD2, dppF2
1804051	1808107	+	4	ddpA, Hsero_1569, Hsero_1570, Hsero_1571
1811697	1813122	+	2	Hsero_1578, Hsero_1579
1814097	1814402	+	2	Hsero_1581, Hsero_1582
1814452	1816725	-	2	Hsero_1583, eriC
1816826	1817994	+	2	Hsero_1585, ssuA
1823426	1823766	-	2	Hsero_1594, Hsero_1595
1826527	1829297	-	2	copA, copB

Inicio	Fim	Strand	Nº de genes	Genes
1829390	1832219	-	2	Hsero_1603, zntA
1833793	1836457	-	2	soxR, copA
1836951	1839626	-	3	Hsero_1610, rimL, moxR
1839846	1842274	+	2	Hsero_1614, Hsero_1615
1842429	1843117	+	2	Hsero_1616, Hsero_1617
1845024	1846393	-	2	nth, fdx
1846523	1847921	-	2	phaZ, Hsero_1623
1855428	1856060	-	2	acb, fkpA
1856370	1858206	+	2	Hsero_1634, Hsero_1635
1858303	1860131	+	2	Hsero_1636, Hsero_1637
1862054	1863945	+	2	Hsero_1640, Hsero_1641
1870471	1874353	+	4	ssuA, ssuD, ssuC, ssuB
1874694	1876139	+	2	Hsero_1653, Hsero_1654
1876303	1880115	+	4	cysU, cysW, cysA, cysB
1885834	1888537	+	3	livM, livG, livF
1888765	1892607	+	3	Hsero_1667, Hsero_1668, livF
1892829	1896309	+	3	Hsero_1670, Hsero_1671, Hsero_1672
1901916	1903474	+	2	Hsero_1679, alkB
1922605	1931540	+	7	Hsero_1698, cheW, cheB, Hsero_1701, cheR, cheY, cheA
1936203	1939572	+	2	Hsero_1708, Hsero_1709
1944628	1947338	+	3	opuBC, opuBB2, opuA
1949471	1951549	-	2	Hsero_1719, rhIE
1951717	1953020	-	2	Hsero_1721, Hsero_1722
1954230	1955781	+	2	lasT, rsuA
1956412	1957886	-	2	Hsero_1726, Hsero_1727
1961222	1963090	-	3	Hsero_1732, Hsero_1733, fecI
1963637	1965848	+	2	ilvI, ilvH
1971086	1973310	-	2	Hsero_1743, Hsero_1744
1977113	1979351	-	2	Hsero_1747, dsd
1979520	1981896	+	2	Hsero_1749, ndeD
1990244	1991952	-	2	Hsero_1758, Hsero_1759
1992222	1994653	-	3	Hsero_1760, Hsero_1761, Hsero_1762
1994910	1996436	+	2	Hsero_1763, qor
1998086	2000784	+	4	nuoA, nuoB, nuoC, nuoD
2000892	2002663	+	2	nuoE, nuoF
2002746	2006647	+	3	nuoG, nuoH, nuoI
2006788	2007774	+	2	nuoJ, nuoK
2007890	2011493	+	2	nuoL, nuoM
2011652	2013487	+	2	nuoN, Hsero_1781
2018270	2021863	+	3	livH, livG, livF
2028992	2030084	-	2	Hsero_1799, Hsero_1800

Início	Fim	Strand	Nº de genes	Genes
2032590	2036388	+	3	Hsero_1802, Hsero_1803, Hsero_1804
2036513	2039914	+	2	Hsero_1805, Hsero_1806
2049691	2051342	+	2	ilvA, rhtB
2053026	2054573	+	2	Hsero_1813, Hsero_1814
2054690	2055458	-	2	Hsero_1815, Hsero_1816
2058296	2059954	-	3	Hsero_1820, Hsero_1821, Hsero_1822
2071679	2078130	+	4	Hsero_1829, Hsero_1830, yhiH, yhhJ
2078428	2081531	-	2	parC, Hsero_1834
2089197	2093042	+	3	Hsero_1840, Hsero_1841, Hsero_1842
2094884	2098642	+	4	dppB, dppC, dmpA, dppA
2098744	2099678	+	2	Hsero_1848, Hsero_1849
2104379	2105603	-	2	mog, Hsero_1856
2115996	2117306	-	2	Hsero_1865, ppiB2
2117436	2119201	-	2	ppiB1, Hsero_1868
2121108	2123160	+	2	cysS, alkA
2134055	2137159	-	3	moeA, mobA, moaA
2142567	2143610	+	2	gph, Hsero_1888
2159112	2160352	-	2	Hsero_1901, trxB
2179539	2181640	+	2	plsX, fabH
2181714	2183398	+	2	fabD, fabG
2183562	2185076	+	2	acpP, fabF1
2186432	2189641	+	3	rseA, rseB, degQ
2190206	2192915	+	2	lepA, lepB
2193109	2194994	+	2	rnc, era
2196105	2200192	+	4	pdxJ, acpS, Hsero_1941, uvrC
2201625	2208021	-	4	Hsero_1948, Hsero_1949, Hsero_1950, tolC
2208252	2210158	-	2	Hsero_1952, Hsero_1953
2210391	2214036	+	3	Hsero_1954, Hsero_1955, Hsero_1956
2215247	2217749	+	3	cobT, cobS, gpmB
2217756	2219808	-	2	Hsero_1961, Hsero_1962
2220043	2222347	+	2	cstA, Hsero_1964
2229810	2232416	+	2	yjgF, relA
2237074	2237659	+	2	rpmI, rplT
2241549	2242274	+	2	ihfA, Hsero_1984
2250071	2252618	+	2	epsG, epsH
2252662	2255381	+	2	epsM, Hsero_1995
2265440	2269117	+	3	ugd, Hsero_2004, wzxC
2269173	2271019	+	2	epsJ, Hsero_2007
2272287	2273881	+	2	cysE, galE
2281390	2282666	-	2	czcD, Hsero_2018
2284732	2286206	+	2	cheR, cheD
2287520	2288567	+	2	cheY, cheZ

Inicio	Fim	Strand	Nº de genes	Genes
2288842	2295193	+	5	flhB, flhA, flhF, fleN, fliA
2295283	2296131	-	2	flgN, Hsero_2031
2297196	2300106	+	4	flgB, flgC, flgD, flgE
2300221	2301778	+	2	flgF, flgG
2301844	2303672	+	2	flgH, flgI
2304879	2308370	+	2	flgK, flgL
2308468	2313082	-	3	Hsero_2044, Hsero_2045, Hsero_2046
2313193	2315061	-	3	fliR, fliQ, fliP
2315108	2317755	-	4	fliO, fliN, fliM, fliL
2319235	2321859	-	3	fliJ, fliI, fliH
2321913	2324815	-	3	fliG, Hsero_2059, Hsero_2060
2325410	2327568	-	3	Hsero_2062, flhB, Hsero_2064
2327706	2330606	-	3	fliT, fliS, fliD
2332419	2334669	-	2	Hsero_2071, Hsero_2072
2335914	2336600	+	2	rpsF, priB
2336656	2337419	+	2	rpsR, rplI
2339124	2339901	+	2	Hsero_2079, Hsero_2080
2344929	2347884	-	3	Hsero_2085, fmt, pmrF
2349308	2351394	-	2	Hsero_2089, arnT
2351990	2353460	+	2	Hsero_2092, aspC
2357288	2359496	+	3	moeA, moaD, Hsero_2099
2363007	2364584	+	2	Hsero_2103, Hsero_2104
2365869	2369533	-	2	Hsero_2106, Hsero_2107
2369742	2371119	+	2	Hsero_2108, crcB
2374338	2376544	+	2	Hsero_2111, menG
2380684	2382843	-	3	modC1, modB1, modA1
2387815	2389312	-	2	Hsero_2122, mgsA
2393546	2396046	-	2	spoT, rpoZ
2398951	2400782	+	2	Hsero_2134, hemN
2400925	2404181	+	2	Hsero_2136, Hsero_2137
2410079	2413353	+	4	Hsero_2145, Hsero_2146, Hsero_2147, Hsero_2148
2417464	2420177	-	3	Hsero_2152, risS, risA
2424525	2426934	-	3	hemK, dapE, arsC
2435464	2438879	+	2	Hsero_2166, lig
2445761	2448817	+	3	Hsero_2172, mdlB, Hsero_2174
2457269	2461432	+	4	uppS, cdsA, dxr, Hsero_2185
2464654	2466199	+	2	lpxD, fabZ
2466412	2468359	+	2	lpxA, lpxB
2468428	2469862	+	2	rnhB, spoU
2474597	2476870	+	2	Hsero_2197, Hsero_2198
2476968	2479080	-	2	Hsero_2199, Hsero_2200
2481554	2482283	+	2	Hsero_2203, Hsero_2204

Inicio	Fim	Strand	Nº de genes	Genes
2490362	2492366	+	3	Hsero_2211, Hsero_2212, ldcA
2493205	2496882	+	3	uup, nadA, nadC
2498263	2502148	-	2	Hsero_2219, nadB
2503899	2504566	+	2	Hsero_2222, Hsero_2223
2504591	2506250	-	2	Hsero_2224, Hsero_2225
2512412	2516357	-	3	Hsero_2230, Hsero_2231, Hsero_2232
2521991	2524457	-	2	Hsero_2238, Hsero_2239
2533050	2533858	-	2	Hsero_2244, ispZ
2536058	2539668	+	4	livH, livM, livG, livF
2547811	2552923	-	3	dnaE, Hsero_2259, Hsero_2260
2560353	2562389	-	2	Hsero_2268, poxB
2562602	2566218	+	2	Hsero_2270, Hsero_2271
2568938	2570769	-	2	Hsero_2274, Hsero_2275
2570825	2576774	-	4	Hsero_2276, Hsero_2277, Hsero_2278, Hsero_2279
2579718	2583526	-	2	Hsero_2282, acnA
2586315	2587753	+	2	glnP, glnQ
2595478	2598819	+	4	Hsero_2295, Hsero_2296, Hsero_2297, Hsero_2298
2609597	2615211	-	6	Hsero_2307, Hsero_2308, Hsero_2309, coxB, coxA, coxB
2615344	2617465	-	4	Hsero_2313, Hsero_2314, Hsero_2315, Hsero_2316
2618905	2623053	+	2	Hsero_2319, Hsero_2320
2624099	2626740	+	3	Hsero_2322, Hsero_2323, Hsero_2324
2627048	2636403	+	3	Hsero_2325, Hsero_2326, Hsero_2327
2637100	2638763	-	2	tauC, tauB
2640109	2642952	-	2	csdB, Hsero_2333
2643030	2644280	-	2	Hsero_2334, cysE
2648532	2651225	+	4	Hsero_2339, Hsero_2340, Hsero_2341, Hsero_2342
2682281	2684638	-	2	Hsero_2346, Hsero_2347
2690246	2691813	+	2	Hsero_2353, Hsero_2354
2695932	2696646	+	2	Hsero_2361, Hsero_2362
2699585	2702021	+	5	Hsero_2367, Hsero_2368, Hsero_2369, Hsero_2370, Hsero_2371
2702453	2703092	-	2	Hsero_2372, Hsero_2373
2718734	2723431	+	5	Hsero_2385, Hsero_2386, Hsero_2387, Hsero_2388, Hsero_2389
2723460	2724171	-	2	Hsero_2390, Hsero_2391
2727385	2728318	+	2	Hsero_2397, Hsero_2398
2732028	2734898	+	3	dksA, phbC, fabI
2736489	2737701	-	2	Hsero_2408, Hsero_2409
2742290	2743293	+	2	Hsero_2414, Hsero_2415
2743328	2745238	-	2	Hsero_2416, Hsero_2417
2745286	2746518	-	2	Hsero_2418, rpoE
2747357	2749371	+	2	Hsero_2421, Hsero_2422

Inicio	Fim	Strand	Nº de genes	Genes
2751737	2757531	+	4	Hsero_2425, Hsero_2426, iolC, iolD
2757592	2759303	+	2	iolE, iolB
2761439	2763015	-	2	fecR, fecI
2763406	2765384	+	2	Hsero_2435, baeS
2778102	2784173	+	6	bioB, bioI, bioF, bioA, bioD, Hsero_2452
2786015	2789357	-	3	livF, livG, livH
2791393	2796730	-	5	hydH, shb1, cyoD, cyoC, cyoB
2805227	2807592	-	3	Hsero_2471, Hsero_2472, Hsero_2473
2811384	2814493	+	2	cls, Hsero_2477
2815243	2816670	-	2	Hsero_2479, lepB
2818487	2822850	+	2	Hsero_2482, Hsero_2483
2822943	2825055	+	2	Hsero_2484, Hsero_2485
2825086	2826458	-	2	serA, Hsero_2487
2837739	2839849	-	2	Hsero_2494, Hsero_2495
2843031	2854770	-	5	treY, malQ, treZ, glgX, glgB
2864673	2865380	-	2	Hsero_2513, Hsero_2514
2868243	2870515	-	2	eutC, eutB
2873546	2875676	+	2	eutR, Hsero_2521
2879259	2880940	-	2	tauC, tauB
2881114	2883239	+	2	ygiX, ygiY
2888962	2891278	-	2	Hsero_2535, Hsero_2536
2892990	2895646	+	3	Hsero_2539, Hsero_2540, cpdA
2897704	2899832	+	2	Hsero_2544, Hsero_2545
2900896	2901842	-	2	Hsero_2547, Hsero_2548
2902001	2903962	-	2	caiD, caiA
2911275	2915956	+	3	Hsero_2555, Hsero_2556, gst
2919596	2921746	-	2	paaK, Hsero_2562
2930274	2934286	-	3	Hsero_2570, Hsero_2571, Hsero_2572
2935403	2938936	+	3	papD, papC, Hsero_2576
2955951	2957079	+	2	Hsero_2589, crtB1
2958342	2960522	+	2	crtB, hpnE
2961977	2963424	-	2	Hsero_2595, Hsero_2596
2963484	2965399	-	2	Hsero_2597, Hsero_2598
2967838	2969494	-	2	holB, tmk
2971803	2972923	+	2	Hsero_2604, Hsero_2605
2974192	2978159	-	4	plcN, tonB, exbD, Hsero_2610
2981463	2985713	+	4	purR, ugpB, ugpA, ugpE
2991321	2992068	-	2	Hsero_2621, Hsero_2622
2996638	2997756	+	2	Hsero_2626, Hsero_2627
2999536	3003281	-	4	amiE, Hsero_2631, Hsero_2632, Hsero_2633
3006228	3010599	+	2	Hsero_2636, Hsero_2637
3015683	3025017	-	9	cobQ, cobC, cobD, cobP, nfnB, cobB, cobA,

Inicio	Fim	Strand	Nº de genes	Genes
				Hsero_2648, cbiJH
3025161	3037183	-	8	cbiG, cbiF, cbiL, cbiD, cobH, cobL, chlI, cobN
3037565	3038927	-	2	Hsero_2658, Hsero_2659
3056951	3061184	-	5	fabD, mdcB, mdcG, mdcC, accD
3061531	3065986	-	4	mdcA, dctM, dctQ, dctP
3066969	3069504	-	3	rluA, Hsero_2689, Hsero_2690
3070092	3070794	+	2	Hsero_2692, Hsero_2693
3071245	3072569	-	2	Hsero_2695, rpoE
3073140	3074611	-	2	Hsero_2698, Hsero_2699
3074677	3077715	-	2	Hsero_2700, Hsero_2701
3086406	3088050	-	3	Hsero_2709, Hsero_2710, Hsero_2711
3088310	3090275	-	2	Hsero_2712, adhC
3107692	3116199	-	6	degQ, lysA, caiC, atoC, Hsero_2731, wcaJ
3117527	3119751	+	2	Hsero_2734, Hsero_2735
3119834	3123328	+	3	wecC, galE, Hsero_2738
3125270	3130839	+	5	Hsero_2740, Hsero_2741, Hsero_2742, wecB, Hsero_2744
3130944	3144115	+	10	Hsero_2745, Hsero_2746, Hsero_2747, Hsero_2748, asnB, Hsero_2750, Hsero_2751, Hsero_2752, Hsero_2753, Hsero_2754
3154113	3156454	-	2	Hsero_2763, Hsero_2764
3162662	3165183	-	2	kdsA, pyrG
3166289	3169584	-	2	Hsero_2774, Hsero_2775
3171182	3173150	-	2	lolD, lolE
3175950	3178672	+	2	Hsero_2780, recJ
3194348	3197568	-	3	citE, caiB, caiA
3208650	3214976	-	6	Hsero_2813, tldD, tldE, Hsero_2816, exbD, exbB
3215031	3218208	-	2	Hsero_2819, prhA
3219600	3221066	+	2	prhI, Hsero_2823
3222761	3223996	-	2	Hsero_2825, trxA1
3225255	3226938	+	2	nifS, nifU
3226982	3229603	+	3	Hsero_2831, hesB1, Hsero_2833
3231401	3235333	+	5	nifW, fixA, fixB, fixC, fixX
3235415	3238045	-	3	modC, modB, modA
3238752	3242931	-	6	fdxB, Hsero_2846, Hsero_2847, nifX, nifN, nifE
3248199	3249200	+	2	Hsero_2855, Hsero_2856
3249525	3250717	-	2	Hsero_2858, fdx
3250791	3252440	-	3	Hsero_2860, Hsero_2861, fixU
3252485	3257563	-	8	nifS2, Hsero_2864, nifZ1, nifZ, Hsero_2867, hesB, fdxN, nifB
3260788	3262595	+	2	Hsero_2873, modE1
3271086	3276050	-	3	tolC, hlyD, cyaB
3281384	3284884	+	4	Hsero_2884, Hsero_2885, Hsero_2886, Hsero_2887

Inicio	Fim	Strand	Nº de genes	Genes
3285961	3287066	-	2	lrgB, Hsero_2890
3287279	3288671	+	2	Hsero_2891, Hsero_2892
3289537	3290414	+	2	Hsero_2894, Hsero_2895
3290763	3294655	+	2	narK, nirB
3294802	3296407	+	2	nirD, Hsero_2899
3299680	3300994	-	2	Hsero_2903, Hsero_2904
3312723	3313495	-	2	Hsero_2916, Hsero_2917
3318236	3322443	-	4	hemY, hemX, hemD, hemC
3327028	3328299	-	2	Hsero_2927, Hsero_2928
3331440	3334687	+	4	Hsero_2932, Hsero_2933, Hsero_2934, Hsero_2935
3337452	3345535	-	2	pbpC, Hsero_2939
3348617	3351094	-	2	purA, hisZ
3351418	3353663	-	2	hflC, hflK
3353753	3355156	-	2	hflX, hflq
3355296	3357865	-	2	Hsero_2949, Hsero_2950
3360181	3362541	-	2	gcpE, Hsero_2954
3367106	3369121	-	2	rpoS, nlpD
3369349	3374325	-	4	pcm, surE, Hsero_2963, dinG
3374758	3376751	-	2	Hsero_2965, Hsero_2966
3384903	3387403	-	2	sdhB, sdhA
3390696	3392267	+	2	Hsero_2977, Hsero_2978
3396753	3399371	-	2	cheW, cheA
3406500	3408307	-	2	Hsero_2991, Hsero_2992
3408527	3410739	-	3	avrD, mmyX, acpD
3418368	3420239	-	2	Hsero_3002, rluD
3427637	3428455	+	2	Hsero_3010, Hsero_3011
3432528	3441483	-	8	Hsero_3015, Hsero_3016, cheB, cheA, cheW, cheR, Hsero_3021, Hsero_3022
3441733	3443016	-	2	Hsero_3023, serB
3446690	3447954	+	2	ispD, ispF
3451387	3454244	-	2	Hsero_3031, tsr
3460589	3462218	-	3	Hsero_3038, Hsero_3039, Hsero_3040
3496811	3501774	-	3	Hsero_3044, Hsero_3045, sunT
3506790	3508409	-	2	yedZ, Hsero_3051
3508607	3510702	+	2	czcR, czcS
3512279	3515904	+	4	Hsero_3055, Hsero_3056, cydA, cydB
3522282	3524692	-	2	mmsB, mmsA
3530080	3531068	+	2	rpoE, Hsero_3074
3531097	3533156	-	2	Hsero_3075, caiD
3539949	3542010	+	2	Hsero_3085, Hsero_3086
3542072	3545092	+	3	Hsero_3087, Hsero_3088, Hsero_3089
3545715	3546824	+	2	Hsero_3091, Hsero_3092

Inicio	Fim	Strand	Nº de genes	Genes
3549294	3551950	+	2	cheZ, cheA
3554001	3556218	+	2	cysI, Hsero_3100
3557156	3560600	+	4	cysD, cysN, cysG, Hsero_3105
3560679	3563337	-	3	Hsero_3106, Hsero_3107, Hsero_3108
3563489	3566046	+	2	pepA, Hsero_3110
3588260	3589830	-	2	Hsero_3130, Hsero_3131
3589985	3593042	+	2	Hsero_3132, Hsero_3133
3597344	3598501	-	2	rhtB, Hsero_3141
3599547	3601311	-	2	iscS1, Hsero_3145
3611548	3613685	-	2	Hsero_3153, selD
3623486	3626180	-	3	Hsero_3164, Hsero_3165, Hsero_3166
3627250	3630038	+	3	Hsero_3168, Hsero_3169, Hsero_3170
3631153	3632318	-	3	Hsero_3172, Hsero_3173, Hsero_3174
3634222	3640589	-	6	abgB, Hsero_3178, dppF1, dppD1, dppC1, dppB1
3649081	3653314	-	3	oprN, Hsero_3188, emrK
3660600	3662660	-	2	Hsero_3199, fixG
3663942	3665997	-	2	fixO, fixN
3666142	3671155	-	4	fixS, fixI, Hsero_3206, hemN
3675217	3676610	-	2	argE, Hsero_3213
3683314	3684638	+	2	Hsero_3221, aat
3687206	3688263	+	2	chaC, Hsero_3227
3688320	3688869	+	2	Hsero_3228, Hsero_3229
3689181	3691477	-	2	pldB, Hsero_3232
3694526	3696450	-	3	Hsero_3235, rpoE, Hsero_3237
3696723	3700439	+	5	Hsero_3238, Hsero_3239, Hsero_3240, Hsero_3241, Hsero_3242
3703639	3705638	-	2	Hsero_3247, Hsero_3248
3734415	3737894	+	2	xdhB, xdhC
3740253	3744652	+	3	tolC, emrA, emrB
3751761	3753746	-	2	Hsero_3274, Hsero_3275
3765352	3766870	-	2	cvpA, Hsero_3288
3770564	3773450	-	3	trpB, trpF, truA
3788634	3791127	+	3	wrbA, Hsero_3307, Hsero_3308
3809956	3811534	-	2	Hsero_3321, Hsero_3322
3811666	3813437	-	2	Hsero_3323, Hsero_3324
3816032	3817807	+	2	argM, Hsero_3328
3817795	3821723	-	4	Hsero_3329, Hsero_3330, Hsero_3331, Hsero_3332
3829770	3831550	+	2	Hsero_3339, Hsero_3340
3831642	3835485	-	3	Hsero_3341, betA, Hsero_3343
3836758	3838843	+	2	Hsero_3345, Hsero_3346
3842108	3843570	+	2	Hsero_3351, Hsero_3352
3844508	3846242	+	3	Hsero_3354, Hsero_3355, Hsero_3356

Inicio	Fim	Strand	Nº de genes	Genes
3846349	3849488	-	2	dctD, dctB
3851422	3854087	-	2	maeB, Hsero_3361
3854111	3855615	+	2	thiL, pgpA
3857305	3859306	+	3	Hsero_3366, Hsero_3367, Hsero_3368
3868503	3869540	-	2	Hsero_3375, Hsero_3376
3870401	3876466	-	4	Hsero_3377, cheB, cheR, Hsero_3380
3877724	3879782	-	2	Hsero_3382, Hsero_3383
3880139	3883065	+	2	Hsero_3384, piuC
3884659	3886371	-	2	tauC, tauB
3888922	3890970	+	2	Hsero_3391, Hsero_3392
3892648	3898924	+	6	Hsero_3394, Hsero_3395, Hsero_3396, Hsero_3397, Hsero_3398, xylB
3899258	3903885	+	3	Hsero_3401, Hsero_3402, Hsero_3403
3906956	3911277	+	3	Hsero_3406, Hsero_3407, Hsero_3408
3913104	3913778	+	2	Hsero_3410, Hsero_3411
3921502	3925201	-	3	Hsero_3420, Hsero_3421, Hsero_3422
3930659	3934255	-	4	Hsero_3427, pqiB, pqiA, pqiA
3961882	3963074	-	2	chrA, chrA
3982874	3983784	-	2	umuC, umuD
3986292	3989763	-	2	Hsero_3472, spoVK
4007887	4009274	+	2	qor, Hsero_3500
4018726	4019887	-	2	Hsero_3509, Hsero_3510
4020266	4022961	-	3	Hsero_3511, Hsero_3512, glnQ
4023006	4024442	-	2	glnP, glnH
4024890	4029954	-	6	gltL, gltK, gltJ, pcaC, hisJ, Hsero_3521
4038610	4041990	+	2	Hsero_3528, Hsero_3529
4048085	4052132	+	3	fecI, fecR, Hsero_3535
4052413	4055449	+	2	Hsero_3536, fliD
4055493	4057486	+	2	flgE, Hsero_3539
4082297	4085181	+	4	hemF, nadD, Hsero_3561, Hsero_3562
4085237	4087311	+	2	maf, cafA
4088785	4093648	+	4	Hsero_3566, Hsero_3567, rfaJ, rfaJ
4101233	4103052	+	2	Hsero_3575, gltX
4118064	4119043	+	2	Hsero_3590, yjaB
4119120	4121029	-	2	Hsero_3592, Hsero_3593
4121114	4122797	-	3	Hsero_3594, Hsero_3595, Hsero_3596
4123522	4124670	-	2	Hsero_3597, Hsero_3598
4134435	4136485	-	2	nusA, Hsero_3606
4144934	4148142	-	2	Hsero_3613, alaS
4150158	4152124	+	2	Hsero_3618, Hsero_3619
4167290	4168886	+	3	Hsero_3635, Hsero_3636, Hsero_3637
4169190	4171863	+	2	narK1, narU

Início	Fim	Strand	Nº de genes	Genes
4172043	4179734	+	5	narG, narH, narJ, narI, moaA
4179858	4182501	+	2	narX, narL
4182562	4183397	+	2	Hsero_3647, Hsero_3648
4183489	4184885	+	2	Hsero_3649, pilW
4185029	4186450	+	3	Hsero_3651, Hsero_3652, Hsero_3653
4189304	4193835	-	5	Hsero_3657, Hsero_3658, Hsero_3659, Hsero_3660, Hsero_3661
4193951	4196818	-	3	Hsero_3662, nagH, nagG
4212665	4213560	-	2	recR, Hsero_3679
4217968	4220281	-	2	rfaE, ugdH
4220332	4221850	-	2	Hsero_3686, Hsero_3687
4222151	4224227	-	2	ihfB, rpsA
4226552	4228612	-	2	tyrA, Hsero_3693
4240233	4241955	+	3	Hsero_3701, Hsero_3702, Hsero_3703
4242018	4253218	+	11	Hsero_3704, Hsero_3705, Hsero_3706, Hsero_3707, Hsero_3708, hutH, Hsero_3710, Hsero_3711, Hsero_3712, Hsero_3713, fabF
4253685	4255663	+	2	fabG, fabF
4262979	4266357	+	3	Hsero_3724, Hsero_3725, Hsero_3726
4266593	4268925	+	2	Hsero_3727, Hsero_3728
4271794	4274486	+	2	tar, Hsero_3731
4277291	4279998	+	2	Hsero_3733, Hsero_3734
4280100	4285633	+	2	Hsero_3735, Hsero_3736
4286685	4287796	-	2	Hsero_3738, Hsero_3739
4287839	4289667	-	2	Hsero_3740, rsbU
4292565	4296394	-	2	Hsero_3745, Hsero_3746
4297936	4298897	-	2	Hsero_3748, Hsero_3749
4299670	4301540	-	2	Hsero_3751, kaiC
4301624	4303725	-	3	Hsero_3753, Hsero_3754, cheD3
4304069	4307780	+	3	Hsero_3757, Hsero_3758, Hsero_3759
4308065	4308672	+	2	phd, doc
4311680	4314280	-	3	Hsero_3765, aroE, Hsero_3767
4314410	4316030	+	2	pcaR, Hsero_3769
4323470	4327009	-	2	kdpE, kdpD
4327190	4330032	-	2	kdpC, kdpB
4346958	4348938	-	2	upgC, upgE
4351672	4353215	+	2	ubiA, tpm
4353264	4353955	-	2	Hsero_3798, Hsero_3799
4354239	4356078	-	3	Hsero_3800, nlpD, proC
4364147	4368814	-	3	Hsero_3811, Hsero_3812, Hsero_3813
4369087	4370668	-	2	fabG, fabG
4375883	4378574	+	3	fadD, Hsero_3822, Hsero_3823

Inicio	Fim	Strand	Nº de genes	Genes
4380559	4384398	-	3	glcF, glcE, glcD
4390150	4391372	-	2	ispA, xseB
4396865	4401994	-	3	Hsero_3839, pilJ, pilI
4402723	4404885	+	2	thiD, hemL
4405440	4407616	-	3	ccmA, Hsero_3847, argC
4407874	4408704	-	2	rpsI, rplM
4412649	4414097	-	2	gltK, gltJ
4433973	4436034	+	3	hisM, hisM, Hsero_3874
4438700	4440656	+	2	Hsero_3877, fabG
4454035	4457171	+	3	trpE, trpG, trpD
4461550	4463732	-	2	ubiB, Hsero_3901
4466329	4470786	-	2	hit, Hsero_3906
4478324	4480278	-	2	mutY, mutM
4480476	4482907	+	2	Hsero_3916, lolB
4507684	4508974	-	2	cybB, acpD
4517574	4522324	-	3	cheD, Hsero_3944, mhpA
4522641	4524145	-	2	Hsero_3946, Hsero_3947
4524212	4525762	-	2	Hsero_3948, Hsero_3949
4538863	4540762	+	2	Hsero_3958, Hsero_3959
4542168	4543904	-	2	ssb, Hsero_3962
4554244	4556431	+	3	Hsero_3968, Hsero_3969, Hsero_3970
4556501	4557851	+	2	Hsero_3971, Hsero_3972
4560263	4561200	+	2	Hsero_3975, ubiX
4572141	4573365	+	2	pcaH, pcaG
4578251	4580266	-	2	vanB, vanA
4583639	4586982	-	4	livF, livG, livM, livH
4588457	4589769	-	2	pcaJ, pcaI
4595210	4599567	+	4	Hsero_4007, paaJ, Hsero_4009, Hsero_4010
4601066	4604648	+	3	Hsero_4012, Hsero_4013, Hsero_4014
4604707	4607510	+	3	Hsero_4015, aidB, Hsero_4017
4607658	4609568	+	2	Hsero_4018, Hsero_4019
4609602	4613997	-	4	Hsero_4020, Hsero_4021, Hsero_4022, Hsero_4023
4619344	4624940	+	3	Hsero_4027, Hsero_4028, Hsero_4029
4628948	4632333	-	3	Hsero_4032, Hsero_4033, Hsero_4034
4632455	4634985	-	3	Hsero_4035, Hsero_4036, Hsero_4037
4636735	4638857	+	2	Hsero_4040, Hsero_4041
4643638	4646443	-	3	petC, petB, petA
4647995	4650634	+	3	degQ, Hsero_4053, glcC
4656255	4658650	-	4	hisE, hisI, hisF, hisA
4660097	4663341	-	3	hisC, Hsero_4068, hisD
4663386	4665316	-	2	hisG, murA
4665610	4667282	-	2	Hsero_4072, Hsero_4073

Inicio	Fim	Strand	Nº de genes	Genes
4667589	4668552	-	2	Hsero_4074, Hsero_4075
4668768	4669981	-	2	vacJ, Hsero_4077
4670044	4671551	-	2	Hsero_4078, Hsero_4079
4677072	4681934	+	2	Hsero_4085, Hsero_4086
4682828	4684469	+	2	Hsero_4089, Hsero_4090
4695226	4696931	+	2	hipA, Hsero_4100
4699965	4704062	+	3	Hsero_4103, Hsero_4104, Hsero_4105
4705044	4708758	+	2	exuT2, Hsero_4108
4711191	4715495	+	5	Hsero_4111, accB, accC, Hsero_4114, Hsero_4115
4718279	4721938	-	3	livF, livGM, livH
4728335	4729927	+	3	paaB, paaC, paaD
4729991	4735596	+	5	paaE, Hsero_4130, paaG, paaH, paaI
4735646	4739013	+	3	paaK, pacF1, Hsero_4136
4742787	4743980	-	2	kdtB, Hsero_4142
4744027	4747425	-	3	Hsero_4143, Hsero_4144, Hsero_4145
4747817	4750399	+	3	ftsY, ftsE, ftsX
4752953	4755574	-	3	Hsero_4151, ctaB, ctaA
4758370	4759199	-	2	Hsero_4158, coxG
4759360	4762651	-	3	coxA, coxB, Hsero_4162
4764130	4765549	+	2	comFC, cspR
4765611	4767300	-	2	Hsero_4166, Hsero_4167
4776849	4778890	+	2	Hsero_4177, Hsero_4178
4781509	4783196	+	2	Hsero_4180, Hsero_4181
4783761	4786390	+	2	Hsero_4182, pilQ
4786592	4789535	+	2	aroBK, dgt
4798122	4803210	+	5	wecC, galE, Hsero_4195, Hsero_4196, wecE
4803297	4807350	+	3	Hsero_4198, Hsero_4199, asnB
4807398	4813863	+	9	Hsero_4201, mesJ, hisH, hisF, Hsero_4205, lpxD, Hsero_4207, fabH, acpP
4816844	4828107	+	9	asnB, gumH, Hsero_4214, asnB, Hsero_4216, rfaG, rfbD, galT, wecB
4835402	4836058	+	2	Hsero_4228, Hsero_4229
4838683	4841545	-	3	Hsero_4232, hop, Hsero_4234
4841755	4842681	-	2	fadR, Hsero_4236
4845823	4848266	+	2	Hsero_4240, Hsero_4241
4848316	4854468	-	5	Hsero_4242, Hsero_4243, Hsero_4244, Hsero_4245, Hsero_4246
4854537	4861566	-	2	Hsero_4247, Hsero_4248
4869573	4871789	+	2	Hsero_4256, Hsero_4257
4878161	4879752	+	2	Hsero_4262, Hsero_4263
4879939	4883519	+	3	Hsero_4264, Hsero_4265, Hsero_4266
4884833	4885526	-	2	Hsero_4268, Hsero_4269
4886521	4887607	+	2	minD, minE

Inicio	Fim	Strand	Nº de genes	Genes
4891285	4892612	+	2	dsbB, hemK
4898851	4901534	-	3	Hsero_4281, cca, gst
4905309	4906507	-	2	metF, Hsero_4288
4909735	4912363	+	3	htrB, htrB, dapF
4921853	4923399	-	2	Hsero_4303, birA
4943658	4949324	+	4	vgr, Hsero_4322, Hsero_4323, Hsero_4324
4963968	4964926	-	3	Hsero_4337, rnpA, rpmH
4968102	4971572	+	3	Hsero_4341, ycgB, Hsero_4343
4974984	4976089	+	2	Hsero_4347, Hsero_4348
4981098	4982891	-	2	fmt, Hsero_4352
4983786	4986200	-	2	ftsN, argS
4995107	4997551	-	2	atpG, atpA
4997597	4998602	-	2	atpH, atpF
4998998	5000160	-	2	atpB, atpI
5005449	5008203	+	2	Hsero_4374, Hsero_4375
5008274	5009785	+	2	Hsero_4376, dadA
5009960	5013797	-	3	Hsero_4378, Hsero_4379, Hsero_4380
5014046	5016204	+	2	ompR, Hsero_4382
5017063	5019269	-	2	Hsero_4384, Hsero_4385
5021619	5022799	-	3	Hsero_4389, Hsero_4390, Hsero_4391
5028754	5030741	-	2	Hsero_4396, Hsero_4397
5032202	5033382	+	2	fabG, folB
5033485	5034212	+	2	Hsero_4401, Hsero_4402
5034299	5036114	-	2	Hsero_4403, tyrB
5040212	5042795	+	3	galE, rfbB, rfbC
5043316	5045303	+	2	Hsero_4413, wcaG
5045370	5047852	+	3	Hsero_4415, Hsero_4416, Hsero_4417
5048017	5049766	+	2	Hsero_4418, Hsero_4419
5051041	5053058	-	2	Hsero_4421, wcaA
5059892	5063429	-	2	Hsero_4429, Hsero_4430
5064419	5065477	-	2	Hsero_4431, Hsero_4432
5065641	5070593	-	3	Hsero_4433, czcA, czcB
5076754	5082215	-	5	fucK, Hsero_4443, Hsero_4444, Hsero_4445, Hsero_4446
5090398	5095261	-	5	serA, menG, Hsero_4455, Hsero_4456, Hsero_4457
5097193	5098728	+	2	Hsero_4460, menG
5110141	5115176	-	2	gatA, Hsero_4471
5115286	5118258	-	4	Hsero_4472, Hsero_4473, Hsero_4474, Hsero_4475
5122442	5125682	+	2	Hsero_4479, Hsero_4480
5125857	5127343	+	2	fecI, fecR
5130837	5131784	-	2	Hsero_4485, acpD
5136868	5137745	-	2	pinR, Hsero_4491

Inicio	Fim	Strand	Nº de genes	Genes
5137969	5140358	-	2	galM, xylR
5141747	5144486	+	2	xylG, xylH
5144750	5148287	+	3	fabG, Hsero_4498, Hsero_4499
5174019	5175585	-	2	Hsero_4513, fecI
5178208	5180527	-	2	Hsero_4518, Hsero_4519
5181618	5182037	+	2	Hsero_4521, Hsero_4522
5196666	5198797	+	2	ompR, tctE
5208341	5209379	+	2	hemO, Hsero_4545
5212154	5214232	-	2	Hsero_4548, Hsero_4549
5216160	5222207	+	6	Hsero_4551, Hsero_4552, Hsero_4553, nagK, nagA, glmS
5225775	5232792	-	3	Hsero_4559, Hsero_4560, Hsero_4561
5234638	5237693	+	2	betA, Hsero_4565
5242230	5244122	-	2	Hsero_4571, Hsero_4572
5244315	5247668	-	3	Hsero_4573, Hsero_4574, Hsero_4575
5247734	5251556	-	3	nadE, pncB, Hsero_4578
5264507	5266793	+	2	Hsero_4588, toeA
5273347	5275421	-	2	Hsero_4596, Hsero_4597
5276464	5277244	-	2	Hsero_4599, Hsero_4600
5281843	5284104	-	3	Hsero_4605, Hsero_4606, Hsero_4607
5287330	5293276	-	3	Hsero_4610, Hsero_4611, Hsero_4612
5293532	5295690	+	2	baeS, baeR
5297892	5303764	-	3	tolC, Hsero_4617, Hsero_4618
5305048	5306870	-	2	Hsero_4620, Hsero_4621
5306938	5308145	-	2	Hsero_4622, Hsero_4623
5308158	5309507	+	2	Hsero_4624, Hsero_4625
5309513	5311089	-	2	Hsero_4626, Hsero_4627
5316483	5319367	-	2	Hsero_4634, Hsero_4635
5320979	5322338	-	2	atoA, atoD
5332623	5335472	+	2	Hsero_4647, Hsero_4648
5337460	5340828	+	4	livH, livM, livG, livF
5340997	5343794	+	2	Hsero_4654, Hsero_4655
5352569	5356189	+	3	mreC, mreD, mrdA
5359714	5361381	+	2	Hsero_4669, Hsero_4670
5361504	5362752	+	2	gmhA, Hsero_4672
5378219	5379231	+	2	nahD, Hsero_4689
5379309	5383921	+	3	Hsero_4690, accC, Hsero_4692
5401096	5403236	-	2	Hsero_4708, Hsero_4709
5403542	5406700	+	2	Hsero_4710, Hsero_4711
5409498	5413824	+	4	urtB, urtC, urtD, urtE
5415080	5418670	+	5	ureA, ureB, ureC, ureE, ureF
5420023	5424036	-	4	glpD1, Hsero_4727, Hsero_4728, Hsero_4729

Inicio	Fim	Strand	Nº de genes	Genes
5425467	5427559	+	2	Hsero_4731, Hsero_4732
5434673	5438076	-	3	dctD1, Hsero_4740, Hsero_4741
5442747	5443939	-	2	Hsero_4745, folE
5446748	5449059	-	2	Hsero_4751, Hsero_4752
5450146	5452806	-	2	Hsero_4754, Hsero_4755
5468640	5471723	+	2	Hsero_4768, Hsero_4769
5471730	5473490	-	2	Hsero_4770, lipA
5473531	5474648	-	2	lipB, Hsero_4773
5476395	5477580	+	2	Hsero_4776, Hsero_4777
5490855	5492276	-	2	parA, rhtB
5496927	5498804	-	2	Hsero_4792, Hsero_4793
5500333	5501861	-	2	Hsero_4795, Hsero_4796

ANEXO 3 - Tabela dos genes preditos pelo software HGF e expressos com cobertura ≥ 3 .

Os genes em cor azul e com * são os que ainda não estão cadastrados no NCBI.

Gene	Início	Fim	Tam anho	Orientação	Condições
N_1 *	5853	6221	369	<	AD3, AltoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_2 *	22849	23139	291	<	BaixoO2, Nar,
N_3	43614	43865	252	<	AD3, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_4	58008	58412	405	<	AD3, Nitrato, PS1, PS3, PT
N_5	63109	63564	456	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_8 *	91531	92049	519	<	BaixoO2, Controle, Nitrato
N_9 *	160736	161176	441	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_10 *	166119	166412	294	<	Nitrato, PS1, PS3, PT, WR
N_11	166281	166535	255	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_12 *	195716	195868	153	>	AD3, Controle, Nitrato, PT
N_15	203037	203966	930	>	AltoO2, BaixoO2, Controle, Nitrato, PT
N_19	305960	306370	411	>	AltoO2, Controle, Nar, Nitrato, PS3
N_20 *	351015	351185	171	>	AD3, Nar, Nitrato, PS1
N_21 *	357055	357207	153	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_22	405542	405844	303	>	AltoO2, BaixoO2, Controle, Nitrato, PS1, PT
N_23 *	417380	417541	162	<	AD3, AltoO2, BaixoO2, Controle, Nitrato, PS1, PS3, PT
N_24 *	417610	418527	918	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_25 *	437716	437970	255	<	AltoO2
N_26	464682	464882	201	>	AltoO2, Controle, Nar, Nitrato, PS3
N_28 *	499310	499534	225	<	PS3
N_30 *	545545	545823	279	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_31 *	545751	545843	93	<	PS1
N_32	546026	546193	168	<	AD3, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_34	560502	560711	210	<	BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_35	571438	571836	399	<	AltoO2, BaixoO2, Nar, Nitrato
N_36 *	573327	573605	279	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_39	607221	607580	360	<	AD3, Controle, Nitrato, PS1, PS3, PT
N_40 *	612416	612664	249	<	Controle
N_46	614126	614314	189	>	Controle, Nar
N_47	671238	671342	105	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_48 *	677690	678040	351	>	Controle, Nar, Nitrato, PS1, PT
N_50	755475	755663	189	<	AD3, AltoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_51 *	787285	787434	150	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_53	840899	841798	900	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1
N_54 *	854002	854118	117	>	AD1, Nar, Nitrato, PS3

Gene	Início	Fim	Tam anho	Orientação	Condições
N_55 *	855552	855704	153	>	PS3
N_61 *	909224	909343	120	>	Controle, Nitrato
N_62	909586	909642	57	>	Controle, Nar
N_63	941596	942048	453	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_65 *	942587	942799	213	>	PS3
N_66 *	965109	965576	468	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_70 *	974679	974918	240	>	Nitrato
N_71 *	975309	975599	291	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_72 *	975894	976223	330	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_73 *	1012582	1013079	498	>	AltoO2, BaixoO2, Controle, Nar, Nitrato
N_76 *	1025501	1025863	363	>	AD1, BaixoO2, Controle, Nitrato, PS3
N_77 *	1037143	1037310	168	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_78	1082207	1082554	348	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_79	1089009	1089950	942	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_80 *	1121012	1121260	249	>	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato
N_81 *	1181624	1181887	264	>	BaixoO2, Controle, Nar, Nitrato, PT
N_82 *	1181939	1182055	117	<	AltoO2, BaixoO2, Nar, Nitrato, PT
N_83	1189154	1189387	234	>	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_84 *	1208197	1208334	138	>	Nitrato
N_85 *	1236502	1236651	150	>	BaixoO2
N_86 *	1236679	1236861	183	<	AD3, Nar, Nitrato, PS1, PT
N_87 *	1244166	1245071	906	>	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_88 *	1250204	1250533	330	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_89 *	1277466	1277624	159	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_90 *	1283931	1284101	171	<	BaixoO2, Nitrato, PS3
N_92 *	1302505	1302759	255	>	Controle, Nar, Nitrato, PT
N_94 *	1326891	1327037	147	<	AltoO2, BaixoO2, Nar, Nitrato
N_95	1335353	1335541	189	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_96 *	1356868	1357122	255	<	Nitrato
N_98	1364136	1364450	315	>	AD1, BaixoO2, Controle, Nar, Nitrato
N_99	1365361	1365519	159	<	Nar, Nitrato
N_100 *	1372628	1372993	366	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_101	1375511	1376029	519	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_102	1384308	1384538	231	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_103 *	1397334	1397636	303	>	Controle, Nar, Nitrato, PS1, PS3, PT
N_105 *	1405382	1405504	123	>	Nar

Gene	Início	Fim	Tam anho	Orientação	Condições
N_106 *	1418467	1418916	450	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_108 *	1449723	1449980	258	>	Controle, Nar, Nitrato
N_109	1478650	1479987	1338	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_114	1559045	1559254	210	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_115	1569238	1569393	156	>	AD3, Controle, Nar, Nitrato
N_116	1597026	1597196	171	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_118 *	1655791	1656192	402	<	Controle
N_119 *	1689121	1689582	462	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_120 *	1725110	1725301	192	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_121 *	1733919	1734203	285	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_124	1784400	1784657	258	<	AD3, AD1, BaixoO2, Controle, Nar, Nitrato, PT
N_125 *	1788554	1788787	234	<	AD1, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_126	1791024	1791242	219	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_127 *	1810055	1810321	267	>	AltoO2, BaixoO2, Nitrato, PT
N_128	1814615	1814986	372	>	AD3, Controle, Nitrato, PS1, PS3, PT
N_129 *	1820006	1820116	111	>	AltoO2, Nitrato
N_131 *	1867320	1867484	165	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_132 *	1880193	1880375	183	<	Nitrato
N_133	1881824	1882096	273	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT, WR
N_134	1911677	1911961	285	>	AD3, AltoO2, Controle, Nitrato, PS1, PS3, PT
N_135 *	1916502	1916963	462	<	Controle, Nar
N_136 *	1918362	1918505	144	>	Nar, PS1, PS3
N_137 *	1919967	1920188	222	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, WR
N_138 *	1920725	1921036	312	<	BaixoO2, Controle, Nar, Nitrato, PT
N_139 *	1931540	1931785	246	>	Controle, Nar
N_140	1940814	1940990	177	>	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_141	1947293	1947412	120	<	AltoO2, BaixoO2, Controle, Nar, Nitrato
N_142	1955902	1956096	195	<	AD3, Controle, Nitrato, PT
N_143	1956106	1956306	201	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_144 *	1973480	1973848	369	<	AltoO2, BaixoO2, Nar
N_145	1997413	1997772	360	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_146	2022905	2023102	198	<	AD3, AD1, AltoO2, Controle, Nitrato, PS1, PS3, PT, WR
N_147	2026128	2026268	141	>	AltoO2, BaixoO2, Controle, Nitrato, PS1, PT
N_148	2028471	2028923	453	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_149 *	2055077	2055289	213	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_150 *	2060930	2061013	84	>	AD3, Nitrato, PT
N_151 *	2070736	2071221	486	>	BaixoO2, Nitrato

Gene	Início	Fim	Tam anho	Orientação	Condições
N_152	2078188	2078355	168	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3
N_153 *	2087564	2087761	198	>	Controle, Nar, Nitrato
N_154	2128754	2129104	351	>	AltoO2, BaixoO2, Controle, Nar, Nitrato
N_155 *	2140584	2140727	144	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_156	2141065	2141226	162	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_157	2158945	2159277	333	<	BaixoO2, Controle, Nitrato
N_158 *	2171618	2171782	165	>	Nar, Nitrato, PS1, PT, WR
N_159	2177480	2177551	72	<	Nitrato, PT
N_160 *	2179373	2179549	177	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_161 *	2207865	2208107	243	>	BaixoO2
N_162 *	2222391	2222660	270	>	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3
N_163 *	2222667	2223035	369	>	BaixoO2, Controle, Nar, Nitrato
N_164	2233601	2233795	195	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_165 *	2260132	2260248	117	>	AltoO2, Nitrato, PS1, PS3
N_166 *	2271003	2272007	1005	<	BaixoO2, Controle, Nar
N_167 *	2276691	2276930	240	>	AltoO2, Controle, Nar, Nitrato, PT
N_168	2277070	2277924	855	>	AD1, BaixoO2, Controle, Nar, Nitrato, PT
N_169 *	2278057	2278419	363	>	Controle, Nar, Nitrato, PS1, PS3, PT
N_170 *	2278837	2279160	324	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_171	2281088	2281312	225	>	AD3, AltoO2, Nitrato, PT
N_172 *	2284440	2284592	153	<	BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_173	2365679	2365879	201	<	BaixoO2
N_174 *	2378921	2379070	150	>	Nitrato
N_175 *	2389311	2389403	93	>	BaixoO2, Nitrato
N_176	2404986	2406029	1044	>	BaixoO2, Controle, Nitrato
N_177	2420397	2420678	282	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT, WR
N_178 *	2429332	2429580	249	>	AD1, BaixoO2, Nitrato
N_179	2469897	2470373	477	>	AltoO2, BaixoO2, Controle, Nitrato, PT
N_180 *	2471343	2471597	255	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_182	2539760	2540044	285	>	PS3
N_184	2556286	2556522	237	>	AltoO2
N_185	2558014	2558586	573	>	Nitrato
N_186	2558707	2558949	243	>	AD1
N_189 *	2578117	2578437	321	<	AD3, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_190	2579393	2579599	207	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT, WR
N_191 *	2592123	2592302	180	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_192 *	2598877	2599164	288	<	BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT

Gene	Início	Fim	Tam anho	Orientação	Condições
N_194	2617572	2617778	207	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_195 *	2623732	2624166	435	>	Controle, Nitrato, PT
N_196 *	2687406	2687570	165	>	PS3
N_197 *	2693082	2693180	99	>	Controle, Nar
N_198 *	2693184	2693384	201	>	Controle, Nar
N_199	2693381	2693590	210	>	Controle, Nar
N_200 *	2695337	2695501	165	>	Controle, Nar
N_201 *	2697585	2697695	111	>	Controle, Nar
N_202 *	2697703	2697804	102	>	Nar
N_203 *	2698681	2698947	267	>	Controle, Nar
N_205	2699307	2699588	282	>	Nar
N_206 *	2702031	2702192	162	>	Nar, Nitrato
N_207 *	2702192	2702329	138	>	Controle, Nar, Nitrato
N_208 *	2703859	2704053	195	<	Nar, Nitrato, PT
N_209 *	2710255	2710365	111	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT, WR
N_212	2759300	2759617	318	<	PT
N_214	2769540	2769728	189	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_215	2783538	2784146	609	<	BaixoO2
N_216 *	2804402	2804578	177	<	AltoO2, BaixoO2, Controle, Nar, Nitrato
N_217	2811179	2811346	168	>	BaixoO2, Controle, Nar
N_218	2833227	2833601	375	<	BaixoO2, Controle, Nar, Nitrato, PS1, PS3
N_220	2857713	2857895	183	>	BaixoO2, Nitrato
N_221	2883247	2883684	438	<	Nitrato
N_222	2917238	2917393	156	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_223	2919160	2919399	240	>	AltoO2, BaixoO2, Controle, Nitrato
N_224 *	2939812	2940186	375	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_225 *	2948506	2948988	483	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_226 *	2949119	2949556	438	<	AltoO2, BaixoO2, Nitrato, PT
N_227 *	2958269	2958343	75	<	BaixoO2
N_230	3015403	3015600	198	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_231	3037149	3037463	315	<	Controle, Nitrato
N_232	3038754	3038915	162	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_237	3045357	3045608	252	<	Nar, Nitrato
N_238	3046050	3046430	381	>	Controle, Nitrato
N_239	3056260	3056502	243	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT, WR
N_240	3056551	3056799	249	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PT, WR
N_241	3061181	3061495	315	<	AltoO2, BaixoO2, Controle, Nitrato, PT, WR
N_242 *	3147032	3147283	252	>	AltoO2
N_243 *	3153956	3154066	111	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR

Gene	Início	Fim	Tam anho	Orientação	Condições
N_244 *	3156480	3156554	75	<	BaixoO2, Controle, Nar
N_245 *	3169396	3169644	249	>	BaixoO2, Controle, Nitrato
N_246	3183458	3183541	84	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_247 *	3191512	3191877	366	<	AD3, AltoO2, BaixoO2, Nar, Nitrato, PS3, PT
N_248 *	3201408	3201653	246	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_249 *	3207921	3208142	222	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_250 *	3217664	3218035	372	>	Nar, Nitrato
N_251 *	3263078	3263302	225	>	AD1, Controle, Nar, Nitrato, PS3, PT
N_255 *	3286693	3286938	246	>	Controle
N_256 *	3306347	3306697	351	<	BaixoO2, Controle, Nar, Nitrato, PS3
N_257 *	3306645	3306887	243	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1
N_258 *	3308834	3308989	156	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_259	3326261	3326929	669	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_260	3357998	3358642	645	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_261	3374455	3374661	207	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_262	3387406	3387771	366	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_263 *	3402821	3403408	588	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_264	3415743	3416012	270	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_265	3416165	3416425	261	<	AD3, Controle, Nar, Nitrato, PS1, PS3
N_266 *	3424983	3425063	81	>	Controle, Nitrato
N_267 *	3426439	3426564	126	<	AD3, AltoO2, BaixoO2, Nar, Nitrato, PS3, PT
N_268 *	3512034	3512213	180	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_269	3515941	3516111	171	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_270 *	3528069	3528296	228	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, WR
N_271 *	3548712	3548861	150	<	Nitrato, PT
N_272 *	3569487	3569720	234	<	Controle, Nitrato, PT
N_273	3571279	3571539	261	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_274 *	3589685	3589951	267	>	AltoO2, Nar, Nitrato
N_275	3593103	3593336	234	<	AD3, AltoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_276 *	3601396	3601548	153	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_277	3621558	3621812	255	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_278	3626177	3626311	135	<	AD3, AltoO2, PT
N_280 *	3630573	3630674	102	>	PT
N_281	3633428	3633772	345	<	AD3, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_282	3663743	3663937	195	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3,

Gene	Início	Fim	Tam anho	Orientação	Condições
					PT, WR
N_283	3671480	3672160	681	>	BaixoO2, Nitrato
N_286	3688561	3688641	81	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_287	3705635	3705940	306	<	Nar, Nitrato
N_288 *	3707821	3707964	144	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_289 *	3708006	3708434	429	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT, WR
N_290 *	3722115	3722252	138	<	Nitrato
N_291	3732618	3732869	252	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_292	3779873	3780004	132	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_293 *	3791251	3791679	429	>	BaixoO2
N_294	3800185	3800415	231	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_295	3809456	3809725	270	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_296 *	3859311	3859607	297	>	Controle, Nar, PS3
N_297 *	3869895	3870359	465	>	AltoO2, Controle, Nar, Nitrato, PS3, PT
N_299	3877433	3877651	219	<	BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_302 *	3912422	3913180	759	>	Controle, Nar, Nitrato, PT
N_303	3915700	3915933	234	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_304 *	3928239	3928718	480	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_305 *	3991383	3991472	90	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_306 *	3997046	3997300	255	<	AD1, AltoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_307	3997673	3997966	294	<	AD3, AD1, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_308 *	4046496	4046660	165	>	AltoO2, BaixoO2, Controle, Nar, Nitrato
N_309 *	4059457	4059624	168	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_310	4059919	4060041	123	<	AD3, AD1, BaixoO2, Nitrato, PT
N_311 *	4062693	4062998	306	>	Controle, Nar, Nitrato, PS1, PS3, PT
N_312	4075097	4075267	171	<	BaixoO2, Controle, Nar, Nitrato
N_313 *	4079046	4079222	177	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_314 *	4110673	4111008	336	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_315 *	4111929	4112030	102	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_316 *	4116904	4117302	399	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_317 *	4126398	4126490	93	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_318	4126535	4126684	150	>	BaixoO2, Controle, Nitrato, PT
N_319	4148461	4148676	216	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_320 *	4148686	4148829	144	<	AD1, AltoO2, BaixoO2, Controle, Nitrato, PS1, PS3, PT, WR
N_321	4152136	4152276	141	<	AltoO2, Controle, Nar, Nitrato, PT

Gene	Início	Fim	Tam anho	Orientação	Condições
N_322	4155958	4156191	234	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_323	4159098	4159682	585	>	BaixoO2, Controle, Nitrato
N_324	4160018	4160560	543	>	BaixoO2, Controle, Nar, Nitrato, PT
N_325 *	4162953	4163402	450	>	Controle, Nar, Nitrato
N_326 *	4235570	4235659	90	<	Controle
N_327	4253215	4253688	474	>	AltoO2, Nitrato
N_328	4268960	4269334	375	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_329	4274825	4275313	489	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT, WR
N_330 *	4276814	4276981	168	>	AltoO2, BaixoO2, Controle, Nar, PT
N_331	4291735	4292130	396	>	BaixoO2, Controle, Nar
N_332 *	4301706	4302164	459	>	AD1
N_333 *	4316694	4316783	90	<	AD1, AltoO2, Controle, Nar, Nitrato
N_334	4322879	4323451	573	>	BaixoO2, Controle, Nitrato
N_335	4331877	4331966	90	<	Nar
N_337 *	4351270	4351521	252	>	AltoO2, Controle, Nitrato, PT
N_338	4428528	4428992	465	<	BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_339 *	4442507	4442638	132	>	AD3, AD1, AltoO2, Nar, Nitrato, PS1, PS3, PT
N_341	4451537	4451692	156	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_342 *	4471054	4471131	78	>	AD3, BaixoO2, Controle, Nar, Nitrato, PS3
N_343	4471165	4471302	138	>	AD3, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_344	4485213	4485821	609	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_348 *	4530542	4530763	222	<	Nitrato, PS3
N_349 *	4535378	4535647	270	<	AD3, AltoO2, Controle, Nar, Nitrato, PS1, PS3
N_350 *	4553997	4554179	183	<	AD1, BaixoO2, Controle, Nar, Nitrato, PT
N_351 *	4559857	4559931	75	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_352	4607552	4607782	231	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_353 *	4615085	4615300	216	>	Controle, Nar, Nitrato, PS1, PT, WR
N_354 *	4640784	4641116	333	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT, WR
N_355	4665326	4665562	237	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_356 *	4667411	4667527	117	<	AD3, AD1, AltoO2, BaixoO2, Controle, Nitrato, PS1, PS3, PT
N_357 *	4673112	4673324	213	<	AD3, AltoO2, BaixoO2, Nar, Nitrato, PS1, PS3, PT, WR
N_358 *	4682105	4682206	102	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_359 *	4694881	4695195	315	>	AD1, BaixoO2, Controle, Nar, Nitrato, PT
N_360 *	4698079	4698540	462	>	AltoO2, BaixoO2, Controle, Nar, Nitrato
N_362 *	4739000	4739164	165	<	AD3, Nitrato, PT
N_363 *	4751149	4751745	597	<	AD3, AD1, AltoO2, Controle, Nar, Nitrato, PS3, PT
N_364	4759196	4759306	111	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_365 *	4762473	4762850	378	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_367 *	4769752	4770183	432	>	AltoO2, BaixoO2, Nitrato

Gene	Início	Fim	Tam anho	Orientação	Condições
N_369 *	4790595	4790789	195	<	AltoO2, BaixoO2, Controle, Nitrato
N_370 *	4790759	4791202	444	<	Nitrato
N_371 *	4791315	4791740	426	<	Controle, Nar, Nitrato, PT
N_372 *	4823138	4823518	381	<	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_373 *	4861409	4861579	171	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato
N_374 *	4866171	4866419	249	>	AD3, AD1, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_375 *	4873236	4873319	84	<	AltoO2, Controle, Nar, Nitrato, PS1
N_376 *	4876754	4876900	147	>	AD1, AltoO2
N_377 *	4932819	4932926	108	<	Nitrato, PS1, PS3
N_378 *	4952646	4952855	210	>	AD3, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_379	4959632	4959712	81	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_380	4976131	4976415	285	<	AD3, BaixoO2, Controle, Nitrato, PS1, PS3, PT
N_382 *	5024166	5024948	783	<	BaixoO2, Controle, Nar, Nitrato
N_383	5030764	5030958	195	<	AD3, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_384 *	5051947	5052054	108	>	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_385 *	5063627	5063827	201	<	Controle, Nar, Nitrato, PS3, PT
N_387 *	5096312	5096509	198	>	AltoO2, BaixoO2, Nar, Nitrato
N_389	5118582	5118734	153	<	BaixoO2, PS1, PS3, WR
N_390 *	5166972	5167166	195	<	BaixoO2, Nitrato, PT
N_391	5188949	5189107	159	>	AltoO2, BaixoO2, Nar, Nitrato
N_392 *	5198810	5199223	414	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_394 *	5225492	5225746	255	>	AD3, AltoO2, Controle, Nar, Nitrato, PS1, PT
N_395 *	5239982	5240134	153	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_396	5240216	5240311	96	<	Nitrato, PS1, PS3, PT
N_398 *	5241041	5241394	354	<	PS1, PT
N_399	5252533	5252757	225	<	AD3, PS1
N_400	5259641	5259928	288	<	AD3, Controle, Nar, Nitrato, PS1, PS3, PT
N_401	5264092	5264271	180	<	Nitrato
N_402 *	5277423	5277863	441	<	Nitrato, PT
N_403	5308167	5308577	411	<	Controle, Nitrato, PT
N_404 *	5328032	5328229	198	>	AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_405 *	5330805	5330948	144	>	BaixoO2, Controle, Nitrato, PT
N_406 *	5332442	5332624	183	<	AltoO2, BaixoO2, Nitrato, PS1, PT
N_407 *	5335763	5335939	177	>	PT
N_408 *	5352403	5352621	219	>	AD3, AD1, AltoO2, Controle, Nitrato, PS1, PS3, PT
N_409 *	5368697	5368936	240	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT
N_410 *	5369590	5369847	258	>	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS3, PT
N_411	5385613	5385786	174	<	AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PT
N_412 *	5385885	5386091	207	<	AD3, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR

Gene	Início	Fim	Tam anho	Orientação	Condições
N_413	5387434	5387703	270	>	AltoO2, BaixoO2, Controle, Nar, Nitrato, PT
N_415 *	5399225	5399383	159	<	Controle, PT
N_416 *	5428056	5428325	270	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_417 *	5439484	5439570	87	<	AltoO2, BaixoO2, Nitrato
N_418 *	5441287	5441523	237	<	BaixoO2, Controle, Nar, Nitrato, PT
N_419 *	5485201	5485329	129	>	AD3, AltoO2, Nar, Nitrato, PS3, PT
N_420	5492864	5493106	243	<	AD3, AltoO2, BaixoO2, Nar, Nitrato, PS1, PS3, PT
N_421	5511315	5511548	234	>	AD3, AD1, AltoO2, BaixoO2, Controle, Nar, Nitrato, PS1, PS3, PT, WR
N_422 *	5511706	5511984	279	<	AD3, AltoO2, BaixoO2, Nar, Nitrato, PS1, PS3, PT

ANEXO 4 - tabela dos ncRNA preditos por Moreno (2013) e expressos com cobertura ≥ 5 .

ncRNA	Início	Fim	Tamanho	Sentido	Condições com Cobertura ≥ 5
CUFF.17.51	5000211	5000493	283	<	AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT
CUFF.1692	4834369	4834486	118	<	AD3, Nitrato, PS1, PS3, PT
CUFF.839	2225999	2226140	142	>	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT
Hs_noco_479	328428	328515	90	>	BaixoO2
Hs_noco_516	825901	826300	400	<	BaixoO2
Hs_noco_517	831492	831611	120	<	AltoO2, Nitrato
Hs_noco_519	853978	854161	184	>	AD1
Hs_noco_567	1624918	1624991	74	>	AltoO2
Hs_noco_571	1664205	1664460	256	<	BaixoO2, Nitrato, PT
Hs_noco_584	1880142	1880548	407	<	Nitrato
Hs_noco_631	2579377	2579662	286	>	AltoO2, BaixoO2, Nitrato, PT
Hs_noco_645	3042974	3043384	411	>	PS1
Hs_noco_654	3188002	3188408	407	<	BaixoO2
Hs_noco_662	3290469	3290741	273	>	PS1, PS3
Hs_noco_663	3317231	3317432	202	>	AD3, AltoO2, Nitrato, PS1, PT
Hs_noco_664	3325609	3325783	175	<	Nitrato
Hs_noco_671	3392204	3392349	146	<	PS3
Hs_noco_692	3750872	3751432	561	>	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT, WR
Hs_noco_697	3776244	3776620	377	>	BaixoO2
Hs_noco_712	3989988	3990380	393	>	Nitrato
Hs_noco_713	3995857	3996132	276	>	BaixoO2, Nitrato
Hs_noco_715	4004964	4005107	144	<	Nitrato, PS1, PS3
Hs_noco_716	4010423	4010502	80	>	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT
Hs_noco_731	4232898	4233211	314	<	BaixoO2
Hs_noco_735	4319058	4319138	81	<	Nitrato, PS1, PS3
Hs_noco_736	4321024	4321152	129	>	PS1
Hs_noco_741	4388096	4388200	105	<	AD3, Nitrato, PT
Hs_noco_743	4408988	4409170	183	<	AD3, BaixoO2, Nitrato, PS1, PT
Hs_noco_753	4489489	4489615	127	<	AltoO2, Nitrato, PS3, PT
Hs_noco_758	4565726	4565818	93	>	BaixoO2
Hs_noco_790	5118375	5118838	464	<	PS3
Hs_noco_819	5373859	5374363	505	<	Nitrato, PT
Hs_noco_824	5433895	5434068	174	>	Nitrato
Hs_noco_834	5511199	5511776	578	>	AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT
Hs_small_1.2	4408980	4409076	97	<	AD3, AltoO2, BaixoO2, Nitrato, PS1, PT
Hs_small_16.2	4873548	4873707	160	<	PS1

ncRNA	Início	Fim	Tamanho	Sentido	Condições com Cobertura ≥ 5
Hs_small_827	5475789	5476188	400	<	AD3, AltoO2, BaixoO2, Nitrato, PS1, PT
Hs_sRNA_10.2	4896557	4896712	156	>	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT, WR
Hs_sRNA_27.2	1943525	1943726	202	>	AD1, AD3, BaixoO2, Nitrato, PS3, PT
Hs_sRNA_3.2	3815799	3815921	123	>	Nitrato, PT
Hs_sRNA_32.2	2111171	2111326	156	<	AD3
Hs_sRNA_45.2	2140728	2141053	326	<	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT
Hs_sRNA_47.2	4136551	4136707	157	<	AD1, AD3, AltoO2, Nitrato, PS1, PS3, PT, WR
Hs_sRNA_48.2	2972997	2973111	115	>	AD1, AD3, AltoO2, Nitrato, PT
Hs_sRNA_50.2	551056	551240	185	>	AD1, AD3, BaixoO2, Nitrato, PS1, PS3, PT
Hs_sRNA_52.2	5058503	5058597	95	>	AltoO2
Hs_sRNA_53.2	1383582	1383798	217	<	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT, WR
Hs_sRNA_57.2	417221	417409	189	>	AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT, WR
Hs_sRNA_6.2	2939838	2939938	101	>	PT
Hs_sRNA_8.2	5335529	5335812	284	>	BaixoO2, Nitrato, PT
Hs_sRNA_Gs1	3998440	3998528	89	<	AD1, Nitrato, PS3, PT
Hs_sRNA_Gs11	808938	809200	263	>	AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT, WR
Hs_sRNA_Gs13	4872868	4872949	82	>	AltoO2
Hs_sRNA_Gs14	4873545	4873711	167	<	PS1
Hs_sRNA_Gs2	3990250	3990386	137	>	Nitrato
Hs_sRNA_Gs21	2939764	2940103	340	>	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT, WR
Hs_sRNA_Gs23	935267	935546	280	>	AD1, AD3, AltoO2, Nitrato, PS1, PS3, PT, WR
Hs_sRNA_Gs25	1433087	1433246	160	>	AD1, Nitrato, PS1, PS3, PT
Hs_sRNA_Gs27	5019552	5019706	155	>	AltoO2, BaixoO2, Nitrato, WR
Hs_sRNA_Gs31	1596981	1597171	191	<	AD3, Nitrato, PT
Hs_sRNA_Gs33	2032443	2032547	105	>	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS3, PT
Hs_sRNA_Gs34	2032256	2032442	187	>	AD3, Nitrato, PS1, PS3, PT
Hs_sRNA_Gs4	4127685	4127782	98	<	PT
Hs_sRNA_Gs5	2177255	2177347	93	>	AD3, AltoO2, Nitrato, PS3
Hs_sRNA_Gs7	1915912	1916133	222	<	PT
Hs_sRNA_Gs8	5433936	5434088	153	>	Nitrato
smallRNA_30.2	2973008	2973111	104	>	AD1, AD3, PT
smallRNA_43.2	4564951	4565134	184	<	BaixoO2, PT
smallRNA_44.2	4564991	4565134	144	<	BaixoO2, Nitrato, PT
smallRNA_49.2	4896565	4896712	148	>	AD1, AD3, AltoO2, BaixoO2, Nitrato, PS1, PS3, PT, WR

ncRNA	Início	Fim	Tamanho	Sentido	Condições com Cobertura ≥ 5
sRNA46	1636668	1636831	164	>	AD3, AltoO2, Nitrato, PS3, PT, WR

ANEXO 5: Script em Perl para trimar os arquivos de qualidade

```
#!/usr/bin/perl -w
# Escrito por Vinicius Weiss - UFPR - 2012

$file = shift or die "Usage: meu.pl <fasta>\n";
$out=shift or die "arquivo de saida\n";
$seed=shift or die "\n especifique quanto deseja cortar";
open (AMOSTRA , "<$file");
open OUT, (">$out");
%reads;
while ($line=<AMOSTRA>){
    chomp $line;
    if ($line=~/#/) {

        #print ARQ "$line\n";
        }elseif ($line=~/\>/){
            $id=$line;
        }else{
            $reads{$id}=$line;
        }
    }
}
print "ETAPA 1 --OK\n";
foreach $key (sort keys %reads) {
    $seq= $reads{$key};
    $tam=length($seq) ;
    ($seq)=(substr($seq,0,($tam-$seed+1)));
    print OUT "$key\n$seq\n";
}
```

ANEXO 6: Script em Perl para trimar os arquivos csfasta.

```
#!/usr/bin/perl -w
# Escrito por Vinicius Weiss - UFPR - 2012

$file = shift or die "Usage: meu.pl <fasta>\n";
$out=shift or die "arquivo de saida\n";
$seed=shift or die "\n especifique quantas bases deseja cortar";
open (AMOSTRA , "<$file");
open OUT, (">$out");
%reads;

while ($line=<AMOSTRA>){

    chomp $line;
    if ($line=~/#/) {

        #print ARQ "$line\n";
    }elseif ($line=~/\>/){
        $id=$line;
    }else{
        $reads{$id}=$line;
    }
}
print "ETAPA 1 --OK\n$tam\n";
foreach $key (sort keys %reads){
    $seq= $reads{$key};
    @vecSeq = split(' ', $seq);
    $tam=@vecSeq;
    (@newSeq)=(@vecSeq[0..($tam-$seed)]);
    $seq=join(' ', @newSeq);
    print OUT "$key\n$seq\n";
}
```